

**UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"**

**FACULDADE DE CIÊNCIAS - CAMPUS BAURU**

**DEPARTAMENTO DE COMPUTAÇÃO**

**BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

**NATHAN DEL GRANDE SOUZA**

**ANÁLISE DE DADOS DA SAÚDE PARA MELHORIAS EM  
POLÍTICAS PÚBLICAS**

**BAURU**

**2023**

NATHAN DEL GRANDE SOUZA

## **ANÁLISE DE DADOS DA SAÚDE PARA MELHORIAS EM POLÍTICAS PÚBLICAS**

Trabalho de Conclusão de Curso do Curso  
de Bacharelado em Ciência da Computação  
da Universidade Estadual Paulista “Júlio  
de Mesquita Filho”, Faculdade de Ciências,  
Campus Bauru.

Orientador: Prof. Dr. Higor Amario de Souza

BAURU

2023

S729a Souza, Nathan Del Grande  
Análise de dados da saúde para melhorias em políticas públicas / Nathan Del Grande Souza. -- Bauru, 2023  
44 p. : il., tabs., mapas

Trabalho de conclusão de curso (Bacharelado - Ciência da Computação) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências, Bauru  
Orientador: Higor Amario de Souza

1. Análise de Dados. 2. Geoespacial. 3. Demografia. 4. Saúde. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências, Bauru. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Nathan Del Grande Souza

## **Análise de dados da saúde para melhorias em políticas públicas**

Trabalho de Conclusão de Curso do Curso de Bacharelado em Ciência da Computação da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Faculdade de Ciências, Campus Bauru.

Banca Examinadora

---

**Prof. Dr. Higor Amario de Souza**

Orientador

Universidade Estadual Paulista “Júlio de Mesquita Filho”

Faculdade de Ciências

Departamento de Ciência da Computação

---

**Profa. Dra. Simone das Graças Domingues Prado**

Universidade Estadual Paulista “Júlio de Mesquita Filho”

Faculdade de Ciências

Departamento de Ciência da Computação

---

**Prof. Dr. João Pedro Albino**

Universidade Estadual Paulista “Júlio de Mesquita Filho”

Faculdade de Ciências

Departamento de Ciência da Computação

Bauru, 16 de novembro de 2023.

# Agradecimentos

Em toda essa jornada, agradeço a Deus, que conhece as minhas dificuldades e me guardou em todos os momentos. A minha família, que com muito amor e carinho me deram todo o suporte que eu precisava. Aos meus amigos, quais compartilhei muitas memórias ao longo desses anos. Também agradeço a todos os profissionais que dedicaram e dedicam seu tempo guiando e ensinando os estudantes durante esta etapa importante da vida.

*"E não temais os que matam o corpo, e não podem matar a alma; temei antes aquele que pode fazer perecer no inferno tanto a alma como o corpo."*

Mateus 10:28

# Resumo

O campo da saúde pública enfrenta desafios complexos. Questões de desigualdade na distribuição de médicos e alto deslocamento de pacientes para internações têm implicações profundas na eficiência e equidade do SUS (Sistema Único de Saúde). Este trabalho tem como objetivo investigar esses desafios críticos por meio de análises geoespaciais, gráficas e correlacionais. Visa-se avaliar a distribuição de médicos, o deslocamento de pacientes e a possível relação entre esses fatores. Além disso, busca-se fornecer uma base sólida para melhorar as políticas públicas de saúde no estado de São Paulo. O estudo empregou análises geoespaciais para mapear a distribuição de médicos, gráficos de dispersão para explorar relações entre variáveis-chave, como renda per capita e deslocamento de pacientes, e análises correlacionais com o método de Pearson para quantificar essas relações. As análises revelaram disparidades significativas na distribuição de médicos e desafios relacionados ao deslocamento de pacientes. Correlações importantes foram identificadas, mas a complexidade das variáveis sugere que há fatores adicionais a serem considerados.

**Palavras-chave:** Geoespaciais, gráfico de dispersão, políticas públicas, análise de dados, saúde, variáveis-chave, correlacionais.

# Abstract

The field of public health faces complex challenges. Issues of inequality in the distribution of physicians and the high displacement of patients for hospitalizations have profound implications for the efficiency and equity of the SUS (Unified Health System). This work aims to investigate these critical challenges through geospatial, graphical, and correlational analyses. It seeks to assess the distribution of doctors, patient displacement, and the potential relationship between these factors. Furthermore, it aims to provide a solid foundation for improving public health policies in the state of São Paulo. The study employed geospatial analyses to map the distribution of medics, scatter plots to explore relationships between key variables, such as per capita income and patient displacement, and Pearson correlation analyses to quantify these relationships. The analyses revealed significant disparities in physician distribution and challenges related to patient displacement. Important correlations were identified, but the complexity of the variables suggests that additional factors should be considered.

**Keywords:** Geospatial, scatter plot, public policies, data analysis, health, key variables, correlational.

# Lista de figuras

|                                                                                          |    |
|------------------------------------------------------------------------------------------|----|
| Figura 1 – População no estado de São Paulo . . . . .                                    | 25 |
| Figura 2 – Quantidade de médicos a cada 1000 habitantes por município . . . . .          | 26 |
| Figura 3 – Atendem pelo SUS . . . . .                                                    | 27 |
| Figura 4 – Não Atendem pelo SUS . . . . .                                                | 27 |
| Figura 5 – Quantidade de estabelecimentos a cada 1000 habitantes por município . . . . . | 28 |
| Figura 6 – Médicos x Estabelecimentos . . . . .                                          | 29 |
| Figura 7 – Valor despendido a cada 1000 habitantes por município . . . . .               | 30 |
| Figura 8 – Renda média domiciliar per capita . . . . .                                   | 31 |
| Figura 9 – Médicos x Renda . . . . .                                                     | 32 |
| Figura 10 – Médicos que não atendem via SUS x Renda . . . . .                            | 33 |
| Figura 11 – Balanço de internações . . . . .                                             | 34 |
| Figura 12 – Balanço x Estabelecimentos / 1000 habitantes . . . . .                       | 36 |
| Figura 13 – Balanço x Estabelecimentos totais . . . . .                                  | 37 |
| Figura 14 – Balanço x Hospitais/Pronto-Socorros . . . . .                                | 38 |
| Figura 15 – Balanço x Habitantes . . . . .                                               | 39 |
| Figura 16 – Balanço x IDHM . . . . .                                                     | 40 |
| Figura 17 – Balanço x IDHM - Pop. > 250k . . . . .                                       | 41 |

# Lista de tabelas

|                                                                           |    |
|---------------------------------------------------------------------------|----|
| Tabela 1 – Correlação Estabelecimentos x Médicos . . . . .                | 29 |
| Tabela 2 – Correlação - Médicos x Renda . . . . .                         | 32 |
| Tabela 3 – Correlação - Médicos que não atendem via SUS x Renda . . . . . | 33 |
| Tabela 4 – Correlação - Médicos que atendem via SUS x Renda . . . . .     | 33 |
| Tabela 5 – Correlação - Balanço x Estabelecimentos . . . . .              | 37 |
| Tabela 6 – Correlação - Balanço x Hospitais/Pronto-Socorros . . . . .     | 38 |
| Tabela 7 – Correlação - Balanço x População . . . . .                     | 39 |
| Tabela 8 – Correlação - Balanço x IDH municipal . . . . .                 | 40 |
| Tabela 9 – Correlação - Balanço x IDHM - Pop > 250k . . . . .             | 41 |

# Lista de abreviaturas e siglas

|         |                                                           |
|---------|-----------------------------------------------------------|
| AMS     | Assistência Médico Sanitária                              |
| CNES    | Cadastro Nacional de Estabelecimentos de Saúde            |
| CSV     | Comma-Separated Values                                    |
| IBGE    | Instituto Brasileiro de Geografia e Estatística           |
| IDHM    | Índice de Desenvolvimento Humano Municipal                |
| JSON    | JavaScript Object Notation                                |
| OCDE    | Organização para a Cooperação e Desenvolvimento Econômico |
| SINAN   | Sistema de Informação de Agravos de Notificação           |
| SISCOLO | Sistema de Informação do Câncer do Colo de Útero          |
| SISVAN  | Sistema de Vigilância Alimentar e Nutricional             |
| SUS     | Sistema Único de Saúde                                    |

# Sumário

|          |                                            |           |
|----------|--------------------------------------------|-----------|
| <b>1</b> | <b>INTRODUÇÃO</b>                          | <b>13</b> |
| <b>2</b> | <b>FUNDAMENTAÇÃO</b>                       | <b>14</b> |
| 2.1      | SUS                                        | 14        |
| 2.2      | DataSUS                                    | 14        |
| 2.3      | Tabnet                                     | 15        |
| 2.3.1    | Epidemiológicas e Morbidade                | 15        |
| 2.3.2    | Rede Assistencial                          | 15        |
| 2.3.3    | Demográficas e Socioeconômicas             | 15        |
| 2.4      | Formatos de estruturação de dados          | 16        |
| 2.4.1    | CSV ( <i>Comma-Separated Values</i> )      | 16        |
| 2.4.2    | GeoJSON                                    | 16        |
| 2.4.3    | Códigos municipais                         | 16        |
| 2.4.4    | Tipologia das cidades brasileiras          | 17        |
| 2.5      | Análises gráficas                          | 17        |
| 2.5.1    | Gráfico de dispersão                       | 17        |
| 2.5.2    | Gráfico de pares                           | 18        |
| 2.6      | Análise geoespacial                        | 18        |
| 2.6.1    | Otimização de Intervalos Naturais de Jenks | 18        |
| 2.7      | Métodos de correlação                      | 19        |
| 2.7.1    | Correlação de Pearson                      | 19        |
| <b>3</b> | <b>METODOLOGIA</b>                         | <b>20</b> |
| 3.1      | Descrição dos dados coletados              | 20        |
| 3.1.1    | Dados de saúde                             | 20        |
| 3.1.2    | Demográficos e Socioeconômicos             | 21        |
| 3.1.3    | Dados geoespaciais                         | 22        |
| 3.2      | Padronização e pré-processamento           | 22        |
| 3.3      | Seleção de técnicas de análise             | 22        |
| 3.4      | Análise dos resultados                     | 22        |
| 3.5      | Ferramentas                                | 23        |
| <b>4</b> | <b>RESULTADOS E ANÁLISES</b>               | <b>24</b> |
| 4.1      | Distribuição Populacional                  | 24        |
| 4.2      | Distribuição de médicos                    | 25        |
| 4.3      | Distribuição de estabelecimentos de saúde  | 28        |

|     |                                         |    |
|-----|-----------------------------------------|----|
| 4.4 | Despesa por município . . . . .         | 30 |
| 4.5 | Renda domiciliar . . . . .              | 31 |
| 4.6 | Deslocamento entre municípios . . . . . | 34 |
| 4.7 | Considerações . . . . .                 | 41 |
| 5   | CONCLUSÃO . . . . .                     | 43 |
| 5.1 | Trabalhos Futuros . . . . .             | 43 |
|     | REFERÊNCIAS . . . . .                   | 44 |

# 1 Introdução

O Sistema Único de Saúde é uma das maiores conquistas da sociedade brasileira e representa um marco na história da saúde pública no país. Iniciado com a Constituição Federal de 1988, no artigo 196, é composto pelo Ministério da Saúde, Estados e Municípios, que atuam em conjunto para garantir o pleno funcionamento de um complexo sistema de saúde.

Desde sua criação, tem como principal objetivo garantir acesso universal, integral e gratuito à saúde para toda a população. De acordo com o Ministério da Saúde, os princípios do SUS são: universalização, equidade e integralidade. Segundo a SES-MG (2015), é o único sistema de saúde pública no mundo que atende mais de 190 milhões de pessoas.

Apesar dos avanços conquistados nas últimas décadas, ainda enfrenta desafios importantes para garantir a qualidade e efetividade dos serviços prestados à população. No decorrer deste projeto, dois problemas de destaque foram identificados, desafiando a eficiência e a igualdade no acesso à assistência médica.

O primeiro desafio é a desigualdade na distribuição de médicos. Embora o SUS tenha como premissa a universalização do atendimento, a alocação desigual de médicos em diferentes regiões resulta em disparidades significativas no acesso à assistência médica. Essa desigualdade geográfica na oferta de serviços de saúde desafia a equidade do sistema e impacta diretamente a qualidade do atendimento.

O segundo desafio é o "alto deslocamento de pacientes para serem internados". O SUS visa proporcionar atendimento médico próximo às residências dos pacientes, mas a necessidade de deslocamento excessivo para internações hospitalares tem sido uma realidade enfrentada por muitos brasileiros. Esse fenômeno não apenas gera desconforto para os pacientes, mas também sobrecarrega o sistema de saúde, tornando-o menos eficiente.

Nesse contexto, a análise de dados da saúde se revela como uma ferramenta essencial para a identificação de problemas e a definição de estratégias para a melhoria das políticas públicas de saúde. Com o advento do Departamento de Informática do Sistema Único de Saúde em 1991, a disponibilidade de dados da saúde pública em larga escala se tornou possível, o que abre oportunidades para a realização de análises detalhadas e o desenvolvimento de soluções baseadas em evidências.

Ao longo deste projeto, será investigado esses desafios em profundidade, utilizando dados públicos do DataSUS sobre os municípios do estado de São Paulo. O objetivo é identificar possíveis causalidade dos problemas, visando contribuir para a eficácia e a equidade do SUS e, em última instância, para a saúde e o bem-estar da população.

## 2 Fundamentação

Nesta seção, serão introduzidos conceitos sobre os dados relacionados a saúde que serão estudados, tipos de estruturas de dados, variantes de análises e métodos utilizados durante o decorrer do projeto.

### 2.1 SUS

O SUS (Sistema Único de Saúde) é um sistema público de saúde que atende a toda a população brasileira, independente de renda ou condição social. Foi criado pela Constituição Federal de 1988 e é considerado um dos maiores sistemas de saúde pública do mundo.

É organizado em três níveis de atenção: atenção primária, atenção secundária e atenção terciária. A atenção primária é a porta de entrada do sistema e é responsável pela prestação de serviços básicos de saúde, como consultas médicas, exames e vacinação. A atenção secundária é responsável pela prestação de serviços de média complexidade, como internações e exames especializados. A atenção terciária é responsável pela prestação de serviços de alta complexidade, como cirurgias e tratamentos oncológicos.

O financiamento provém de recursos públicos, provenientes de impostos e contribuições sociais. O orçamento é dividido entre os três níveis de governo (federal, estadual e municipal), (SUS, s.d.).

### 2.2 DataSUS

O DataSUS, departamento de informática do Sistema Único de Saúde do Brasil é uma das instituições mais importantes do Ministério da Saúde, pois é responsável pela coleta, processamento e disseminação de dados de saúde. Criado em 1991 e desde então vem desempenhando um papel fundamental no monitoramento da saúde da população brasileira (DATASUS, s.d.) .

Disponibiliza dados sobre: Indicadores de Saúde e Pactuações, Assistência à Saúde, Epidemiológicas e Morbidade, Rede Assistencial, Estatísticas Vitais, Demográficas e Socioeconômicas, Inquéritos e Pesquisas, Saúde Suplementar (ANS), Informações Financeiras e Estatísticas de acesso ao Tabnet (ver seção 2.3).

A coleta dados é feita de uma variedade de fontes, incluindo sistemas de informação de saúde, sistemas de vigilância epidemiológica e pesquisas de saúde. Esses dados são usados para gerar relatórios, indicadores e mapas que ajudam a identificar problemas de saúde, monitorar a evolução da saúde da população e avaliar o impacto de políticas públicas de saúde.

Os dados estão disponíveis para o público em geral, por meio de seu site. Isso permite que qualquer pessoa acesse informações sobre a saúde da população brasileira e use essas informações

## 2.3 Tabnet

Aplicativo de tabulação desenvolvido pelo DataSUS para facilitar o acesso e a utilização de dados de saúde. O aplicativo permite que os usuários visualizem, tabulem e gerem relatórios de dados de saúde de uma variedade de fontes, incluindo o SUS (Sistema Único de Saúde), o SIM (Sistema de Informações sobre Mortalidade) e o SINASC (Sistema de Informações sobre Nascidos Vivos) <sup>1</sup>.

### 2.3.1 Epidemiológicas e Morbidade

Os dados epidemiológicos são essenciais para entender fatores e apresentar variáveis considerados determinantes na distribuição de doenças em determinado grupo populacional.

Segundo PEREIRA (2007), o conceito de morbidade trata a respeito da incidência e prevalência de uma doença, quantificando portadores em relação a um grupo populacional em determinado momento.

Contém dados provenientes de órgãos como: SINAN (Sistema de Informação de Agravos de Notificação), SISVAN (Sistema de Vigilância Alimentar e Nutricional), SISCOLO (Sistema de Informação do Câncer do Colo de Útero), dentre outros.

### 2.3.2 Rede Assistencial

Disponibiliza dados referente aos recursos humanos, físicos e relativos a estabelecimentos assistenciais, permitindo a análise quantitativa dos tipos de estabelecimentos, região, valor despedido, número e especialização de profissionais dentre outros.

Contém informações provenientes do CNES (Cadastro Nacional de Estabelecimentos de Saúde) e pesquisas AMS (Assistência Médico Sanitária).

### 2.3.3 Demográficas e Socioeconômicas

Exibe informações referentes a população geral, além de fatores econômicos e questões sociais, abrangendo população, trabalho, renda, economia e saneamento. Contém informações provenientes dos Censos realizados pelo IBGE (Instituto Brasileiro de Geografia e Estatística)

Ao decorrer do trabalho, foi utilizado informações relativas a renda domiciliar e IDHM (Índice de Desenvolvimento Humano Municipal) obtidas pelo Censo 2010.

---

<sup>1</sup>Mais informações podem ser obtidas em: <<https://datasus.saude.gov.br/informacoes-de-saude-tabnet/>>

## 2.4 Formatos de estruturação de dados

Esta seção oferece uma visão geral dos tipos de formatos empregados, destacando suas características distintas e relevância para a organização, armazenamento e análise eficaz dos dados.

### 2.4.1 CSV (*Comma-Separated Values*)

O formato CSV é uma maneira eficaz de representar dados tabulares, em que cada linha do arquivo corresponde a um registro e as colunas são separadas por vírgulas. Esse formato é amplamente utilizado para armazenar e compartilhar conjuntos de dados estruturados. O seu uso permite armazenar informações detalhadas sobre os indicadores de saúde, de maneira organizada e acessível. Sua simplicidade o torna uma escolha conveniente para importar, manipular e analisar dados por meio de bibliotecas Python como o Pandas.

### 2.4.2 GeoJSON

O GeoJSON é um formato de codificação aberto projetado para representar dados geoespaciais. Permite armazenar informações geográficas, como coordenadas de pontos, linhas ou polígonos, juntamente com atributos associados a essas geometrias. A capacidade do GeoJSON de incorporar informações geográficas aos dados fornece uma abordagem eficaz para visualizar e analisar tendências geográficas durante a análise.

- Especialmente adequado para representar informações geográficas, permitindo que você associe atributos aos elementos espaciais, como a população de uma área ou as características de um hospital.
- O uso em conjunto com ferramentas como o Folium, detalhado na seção 3.5 permite a criação de mapas interativos que realçam padrões geográficos, auxiliando na compreensão e comunicação de resultados.
- O formato é suportado por várias ferramentas de geoprocessamento e visualização, garantindo interoperabilidade e flexibilidade durante as análises.

### 2.4.3 Códigos municipais

O código dos municípios do IBGE é um código numérico que identifica cada município do Brasil. Ele é composto por 7 dígitos, sendo os dois primeiros dígitos o código da unidade da federação (UF) e os demais referentes ao município dentro da UF. São normalmente utilizados para tratar estatísticas, mapeamento ou questões administrativas

Além dos municípios, os códigos podem ser utilizados para identificar outros tipos de regiões administrativas, como:

- Unidade Federativa.
- Região Geográfica Intermediária.
- Região Geográfica Imediata.
- Mesorregião Geográfica.
- Microrregião Geográfica.

#### 2.4.4 Tipologia das cidades brasileiras

Para quesitos de classificação, é considerado conglomerados urbanos dos municípios com critério unicamente populacional os seguintes tipos:

- **Cidade pequena:** Até 99 mil habitantes.
- **Cidade média:** Entre 99 à 500 mil habitantes.
- **Cidade grande:** Entre 500 mil à 10 milhões de habitantes.
- **Megacidade:** Acima de 10 milhões de habitantes.

## 2.5 Análises gráficas

Esta seção explora as técnicas de análises gráficas utilizadas ao longo do projeto.

### 2.5.1 Gráfico de dispersão

Um gráfico de dispersão é composto por pontos dispostos em um plano cartesiano, em que cada ponto representa um par de valores correspondentes às variáveis em estudo. A disposição dos pontos no gráfico de dispersão oferece informações sobre possíveis relações entre as variáveis. Diferentes padrões podem ser identificados, tais como:

- **Relação Linear Positiva:** Quando os pontos estão dispostos em uma tendência ascendente no gráfico, indica uma correlação positiva entre as variáveis. Nesse contexto, um aumento em uma variável está associado a um aumento na outra.
- **Relação Linear Negativa:** Se os pontos se alinham em uma tendência descendente, sugere uma correlação negativa entre as variáveis. Aqui, um aumento em uma variável está relacionado a uma diminuição na outra.
- **Dispersão Aleatória:** Quando os pontos não seguem uma tendência clara, isso indica uma falta de correlação ou uma relação complexa entre as variáveis.

### 2.5.2 Gráfico de pares

No gráfico de pares, uma matriz de gráficos de dispersão é criada, em que cada célula da matriz representa a relação entre duas variáveis específicas. Isso permite a visualização simultânea de múltiplas correlações e tendências nos dados.

As principais características do gráfico de pares incluem:

- **Diagonal Principal:** A diagonal principal da matriz geralmente exibe histogramas ou gráficos de densidade das variáveis individuais, o que proporciona uma visão da distribuição de cada variável.
- **Células Fora da Diagonal:** Cada célula fora da diagonal principal representa um gráfico de dispersão que ilustra a relação entre duas variáveis diferentes. A disposição dos pontos nesses gráficos de dispersão oferece informações sobre a associação entre as variáveis correspondentes.

## 2.6 Análise geoespacial

A análise geoespacial se baseia na ideia de que a localização é uma dimensão crítica para a compreensão de muitos fenômenos. Ela permite a combinação de dados espaciais, como mapas, com informações tabulares, como dados estatísticos, a fim de revelar padrões, tendências e relações ocultas.

- **Mapeamento:** A criação de mapas é uma parte central da análise geoespacial. Mapas temáticos podem representar diferentes atributos geográficos, como densidade populacional, distribuição de recursos naturais ou padrões de uso do solo.
- **Estatísticas Espaciais:** As estatísticas espaciais são usadas para identificar padrões estatísticos em dados geográficos. Isso inclui a detecção de autocorrelação espacial (a tendência de valores semelhantes ocorrerem próximos uns dos outros) e a análise de agrupamentos de valores semelhantes em uma determinada área.

### 2.6.1 Otimização de Intervalos Naturais de Jenks

O método de otimização de Jenks, também conhecido como quebras naturais (*natural breaks*) ou método de classificação de Jenks, é uma técnica estatística utilizada na análise geoespacial para segmentar dados geográficos em classes ou categorias de maneira a otimizar a variação interna dentro de cada classe e maximizar a diferença entre as classes.

A abordagem de otimização de Jenks é baseada na ideia de que, ao agrupar valores geográficos em classes, é possível identificar padrões e tendências mais claras nos dados. Isso

é particularmente útil quando se lida com dados de intensidade variável, como densidade populacional, índices de desenvolvimento ou valores de recursos naturais.

O método de Jenks é particularmente eficaz para criar categorias significativas em dados geográficos, permitindo uma análise mais precisa e uma representação visual mais informativa. Ele é frequentemente utilizado na cartografia temática para criar mapas de intensidade ou índices que destacam áreas de interesse com base em características geográficas.

## 2.7 Métodos de correlação

### 2.7.1 Correlação de Pearson

A correlação de Pearson, também conhecida como coeficiente de correlação de Pearson ou apenas  $r$ , é uma medida estatística que avalia a força e a direção da relação linear entre duas variáveis quantitativas.

A fórmula da correlação de Pearson é definida pela equação 2.1.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2.1)$$

Onde:

- $r$  representa o coeficiente de correlação de Pearson.
- $x_i$  e  $y_i$  são os valores das duas variáveis em cada observação.
- $\bar{x}$  e  $\bar{y}$  são as médias das respectivas variáveis.

O coeficiente de correlação de Pearson varia de -1 a 1, indicando diferentes graus e direções de correlação:

- $r = 1$  representa uma correlação positiva perfeita, onde as variáveis aumentam juntas linearmente.
- $r = -1$  indica uma correlação negativa perfeita, em que uma variável aumenta à medida que a outra diminui linearmente.
- $r = 0$  sugere ausência de correlação linear entre as variáveis.

## 3 Metodologia

Nesta etapa é introduzido as etapas, métodos , dados e ferramentas utilizadas durante o projeto. A primeira etapa envolveu a pesquisa e estudo das principais técnicas utilizadas na análise de dados, levando em consideração a natureza dos dados que serão investigados.

### 3.1 Descrição dos dados coletados

A coleta foi feita por meio do sistema do DataSUS. O conjunto de dados é disponibilizado de forma tabular em formato CSV, as informações datam do período mais recente disponibilizado. A divisão será feita a partir da origem dos dados, assim como sua natureza.

#### 3.1.1 Dados de saúde

Conforme CNES (2023), o Cadastro Nacional de Estabelecimentos de Saúde atua como um sistema oficial para cadastro de qualquer estabelecimento que preste qualquer tipo de serviço relacionado a saúde no país. É fundamental para o funcionamento de outros sistemas coligados. Além disso, permite o cadastro e disponibilização de informações relativas aos estabelecimentos, sendo a origem principal dos dados coletados a seguir.

##### **CNES - Recursos Humanos**

##### **Quantidade de Médicos especialistas segundo Município**

**Período:** Ago/2023 **Formato:** [619,64]

Quantidade de médicos por município do estado de São Paulo segundo especialização. Os dados tabulares contam com informações de 619 municípios e 62 tipos de especializações diferentes.

##### **Quantidade de Médicos que atendem pelo SUS segundo Município**

**Período:** Ago/2023 **Formato:** [645,04]

Quantidade de médicos por município do estado de São Paulo que atendem ou não pelo SUS. Os dados tabulares contam com informações de 645 municípios.

##### **CNES - Estabelecimentos**

##### **Quantidade de estabelecimentos de saúde por tipo, segundo Município**

**Período:** Ago/2023 **Formato:** [644,39]

Quantidade de estabelecimentos por município do estado de São Paulo segundo categoria. Os dados tabulares contam com informações de 644 municípios e 37 tipos de estabelecimento

diferentes.

### **Recursos Federais do SUS**

#### **Valor despendido por Tipo de Despesa segundo Município**

**Período:** Nov/2006 **Formato:** [645,71]

Valor despendido por município do estado de São Paulo segundo tipo de despesa. Os dados tabulares contam com informações de 645 municípios e 71 tipos de despesa diferentes.

### **Epidemiológicas e morbidade**

#### **Morbidade Hospitalar do SUS - Por local de internação - São Paulo**

**Período:** jul/2023 **Formato:** [281 ,2]

Número de internações por local de internação segundo município do estado de São Paulo. Os dados tabulares contam com informações de 281 municípios.

#### **Morbidade Hospitalar do SUS - Por local de residência - São Paulo**

**Período:** jul/2023 **Formato:** [645,2]

Número de internações por local de residência segundo município do estado de São Paulo. Os dados tabulares contam com informações de 645 municípios.

### **3.1.2 Demográficos e Socioeconômicos**

#### **Renda média domiciliar per capita - São Paulo**

**Período:** 2010 **Formato:** [645,2]

Renda média domiciliar per capita por município do estado de São Paulo. Os dados tabulares contam com informações de 645 municípios.

#### **Índice de Desenvolvimento Humano Municipal - São Paulo**

**Período:** 2010 **Formato:** [645,5]

Índice de desenvolvimento humano municipal do estado de São Paulo. Os dados tabulares contam com informações de 645 municípios e os seguintes índices:

- IDHM
- IDHM Renda
- IDHM Longevidade
- IDHM Educação

### 3.1.3 Dados geoespaciais

#### **Perímetros dos municípios do estado de São Paulo**

A delimitação geográfica dos municípios foi possível ao utilizar dados em formato GeoJSON contendo as seguintes informações:

- Código IBGE
- Nome do município
- Geometria (coordenadas geográficas)

## 3.2 Padronização e pré-processamento

Os dados provenientes do DataSUS são disponibilizados em formato csv, que é um formato de arquivo de texto simples que pode ser facilmente importado para um dataframe do Pandas. No entanto, antes de iniciar a análise dos dados, é necessário realizar uma padronização dos mesmos, a fim de garantir a sua consistência e integridade.

Uma das etapas é a identificação dos municípios. No DataSUS, os municípios são identificados por um código de 6 ou 7 dígitos, que é fornecido pelo IBGE. No entanto, alguns dos dados disponibilizados pelo DataSUS também utilizam o nome do município para a identificação. Para garantir a consistência dos dados foi necessário modificar a identificação dos municípios, utilizando somente o código municipal fornecido pelo IBGE

A seguinte etapa é a alteração dos tipos dos valores fornecidos. Para tornar os dados compatíveis entre si, os tipos das colunas foram alterados, tanto para valores inteiros, decimais e texto.

## 3.3 Seleção de técnicas de análise

Após o levantamento e tratamento de dados, foram selecionadas as técnicas de análise mais adequadas para aplicar aos dados coletados, compatíveis com as características dos dados e as perguntas da pesquisa, como mencionado nas seções 2.5, 2.6 e 2.7. Foram consideradas técnicas de classificação e agrupamento, assim como análises gráficas, geoespaciais e correlacionais, realizando cruzamento com dados de fontes distintas relacionadas ao tema da saúde.

## 3.4 Análise dos resultados

Por fim, os resultados obtidos foram analisados , avaliando a eficácia das técnicas de análise escolhidas para a melhoria das políticas públicas de saúde. Foram identificadas as

limitações do estudo e possíveis direções futuras para a pesquisa.

## 3.5 Ferramentas

Esta seção destaca as ferramentas tecnológicas e metodológicas empregadas neste estudo para a análise e manipulação dos dados.

### **Python**

Python é uma linguagem de programação de alto nível amplamente usada na análise de dados devido à sua simplicidade e versatilidade. Sua vasta gama de bibliotecas e frameworks torna-a ideal para manipulação, visualização e análise estatística de conjuntos de dados. No contexto do projeto, serve como a base para todas as operações de análise, permitindo o desenvolvimento de soluções personalizadas para os desafios específicos da saúde pública.

### **Colaboratory**

O Google Colaboratory ou popularmente "Colab", é um ambiente de desenvolvimento baseado em nuvem que oferece notebooks Jupyter pré-configurados. Ele permite que você escreva e execute código Python, crie visualizações e compartilhe seus notebooks facilmente. A vantagem fundamental do Colab é a capacidade de executar código sem a necessidade de configuração local, particularmente vantajoso para análises colaborativas e de grande escala. Proporciona um ambiente eficiente e acessível para realizar análises complexas em conjunto com as bibliotecas Python.

### **Pandas**

O Pandas é uma biblioteca Python amplamente usada para análise de dados e manipulação de tabelas. Oferece estruturas de dados que permitem carregar, limpar, transformar e explorar dados de maneira eficaz. No âmbito do estudo, desempenha um papel crucial na preparação dos dados para análise, permitindo a organização e transformação dos dados brutos em formatos mais úteis e compreensíveis.

### **Folium**

Folium é uma biblioteca Python voltada para visualização de dados geoespaciais em mapas interativos. É possível usar o Folium para criar visualizações geográficas das análises. Isso permite a representação espacial das informações, destacando padrões e disparidades regionais de maneira clara, auxiliando na comunicação eficaz dos resultados encontrados.

### **Seaborn**

O Seaborn é uma biblioteca de visualização de dados construída sobre o Matplotlib, outra biblioteca amplamente utilizada para gráficos em Python. O que torna o Seaborn notável é sua capacidade de simplificar a criação de visualizações estatísticas complexas, oferecendo uma interface de alto nível para a construção de gráficos informativos e esteticamente agradáveis.

## 4 Resultados e Análises

Neste capítulo, apresentam-se os resultados e análises derivados do estudo dos dados coletados. Os resultados são fundamentados em análises geoespaciais, as quais se refletem em mapas temáticos gerados utilizando o método otimização de intervalos naturais de Jenks, detalhado na seção 2.6.1, análises gráficas realizadas por meio de gráficos de dispersão e análises correlacionais matemáticas utilizando o método de Pearson, explicado na seção 2.7.1. Essas abordagens forneceram visões relativas à distribuição de profissionais médicos e aos deslocamentos de pacientes, e, consequentemente, demonstram os desafios e oportunidades inerentes à eficiência e à equidade do SUS no contexto do estado de São Paulo.

Ademais, cabe salientar que todas as análises, códigos e *scripts* utilizados neste estudo estão disponíveis de forma aberta e acessível<sup>1</sup>. Esta abordagem visa promover a transparência e replicabilidade da pesquisa, permitindo que outros examinem e construam sobre o nosso trabalho. Encorajando o uso e colaboração com o código e os resultados apresentados, a fim de promover avanços contínuos nas políticas de saúde pública.

### 4.1 Distribuição Populacional

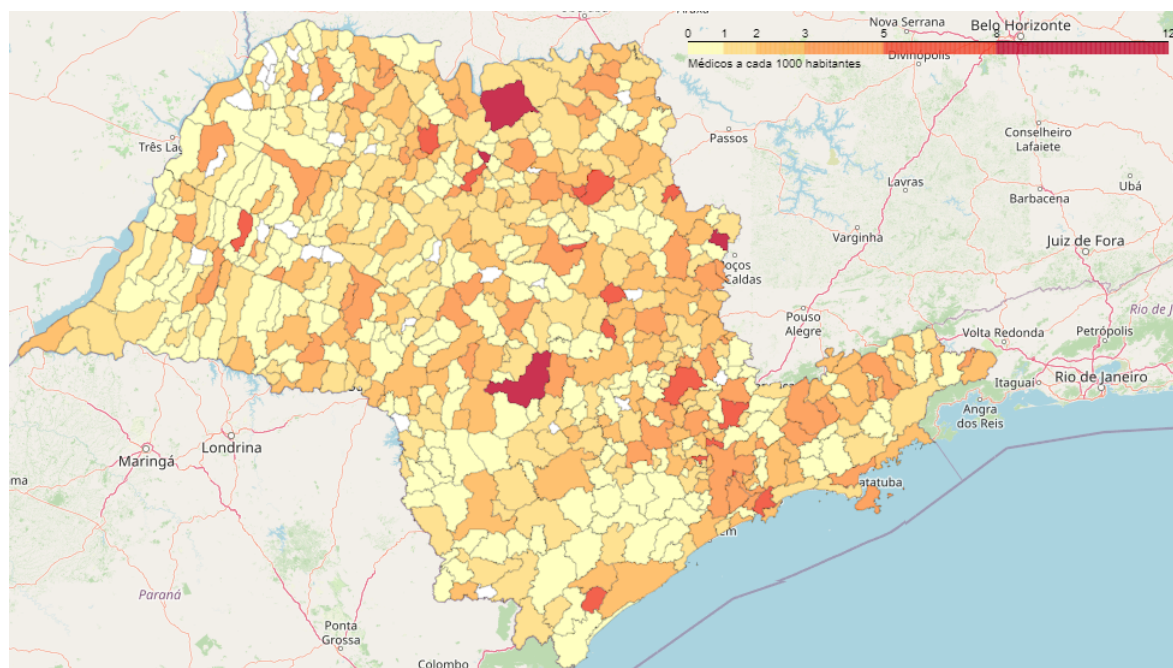
Antes de avaliar os principais resultados, é importante conhecer a demografia regional do estado de São Paulo, que pode ser vista na Figura 1. Composto por 645 municípios, apenas 33 situam-se acima da marca de 250 mil habitantes, dentre esses, 9 possuem mais que 500 mil habitantes, com destaque para a capital, que segundo as últimas estimativas do censo demográfico de 2022 alcançou 11.451.245 habitantes.

---

<sup>1</sup><<https://github.com/Aerosouza/AnaliseDadosSaude>>



Figura 2 – Quantidade de médicos a cada 1000 habitantes por município



Fonte: Elaborada pelo autor.

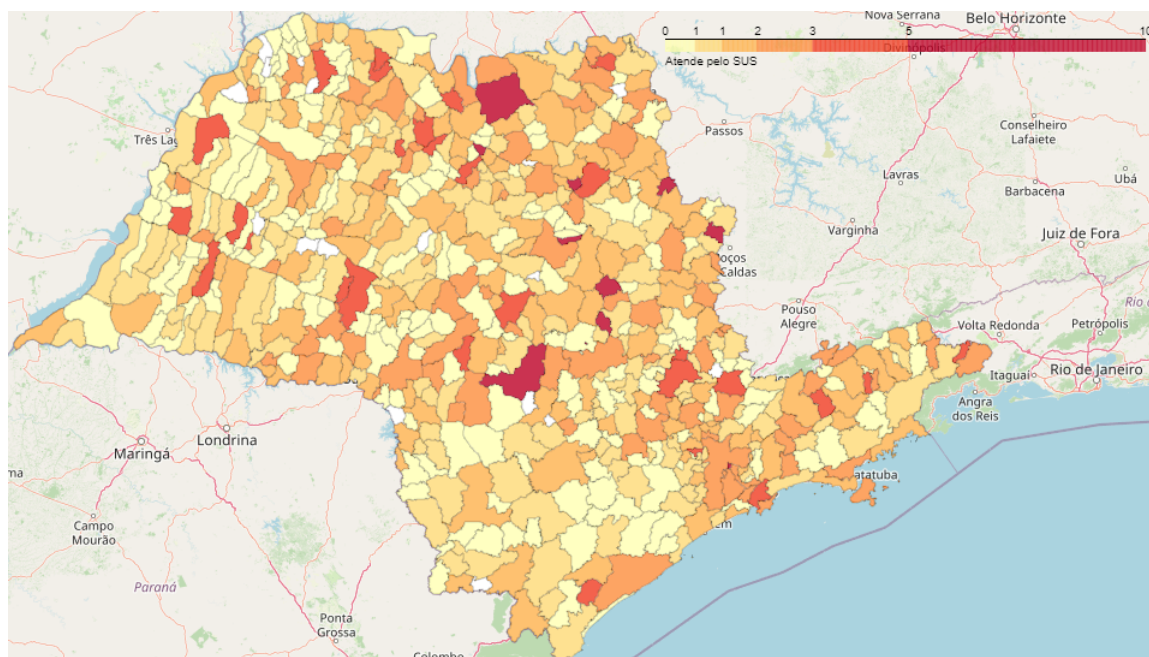
Os destaques ficam para os municípios de Divinolândia, Botucatu e Barretos, com taxas de 12,01, 9,52 e 8,49 respectivamente, únicos com valores superiores a 8, destacados com tom de vermelho mais escuro.

Entre os 10 municípios com as maiores taxas, 5 são caracterizados como cidade pequena, 4 são considerados cidades médias e apenas a cidade de Campinas representa uma cidade grande, com uma taxa de 6,48. A média populacional é de 204.813 habitantes.

As Figuras 3 e 4 exploram as diferenças entre a quantidade de médicos que atendem ou não via SUS a cada 1000 habitantes.

Dentre os resultados, é possível notar, além da notável diferença espacial, que, a maior parte dos profissionais atua exclusivamente ou parcialmente via sistema público. Outro ponto a ser considerado é a extensão territorial atendida, segundo os dados disponibilizados, os municípios representados em cor branca na Figura 4 indicam a ausência de profissionais que não atendem pelo SUS; Em contrapartida, na Figura 3, a presença de médicos abrange todo o território estadual.

Figura 3 – Atendem pelo SUS

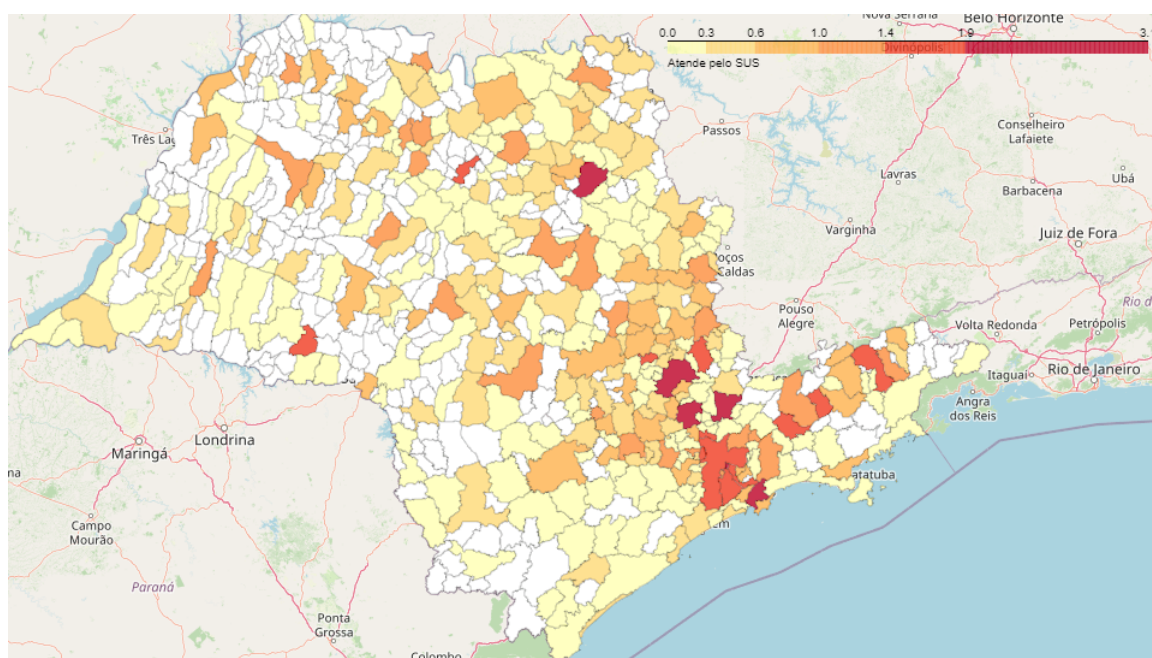


Fonte: Elaborada pelo autor.

Os três municípios com as maiores taxas de profissionais que atendem pelo SUS são: Divinolândia, Embaúba e Botucatu, com taxas de 10,22, 9,04 e 8,01 respectivamente.

Entre os 10 municípios com as maiores taxas, 8 são caracterizados como cidades pequenas e 2 são considerados cidades médias. A média populacional é de 34.061 habitantes.

Figura 4 – Não Atendem pelo SUS



Fonte: Elaborada pelo autor.

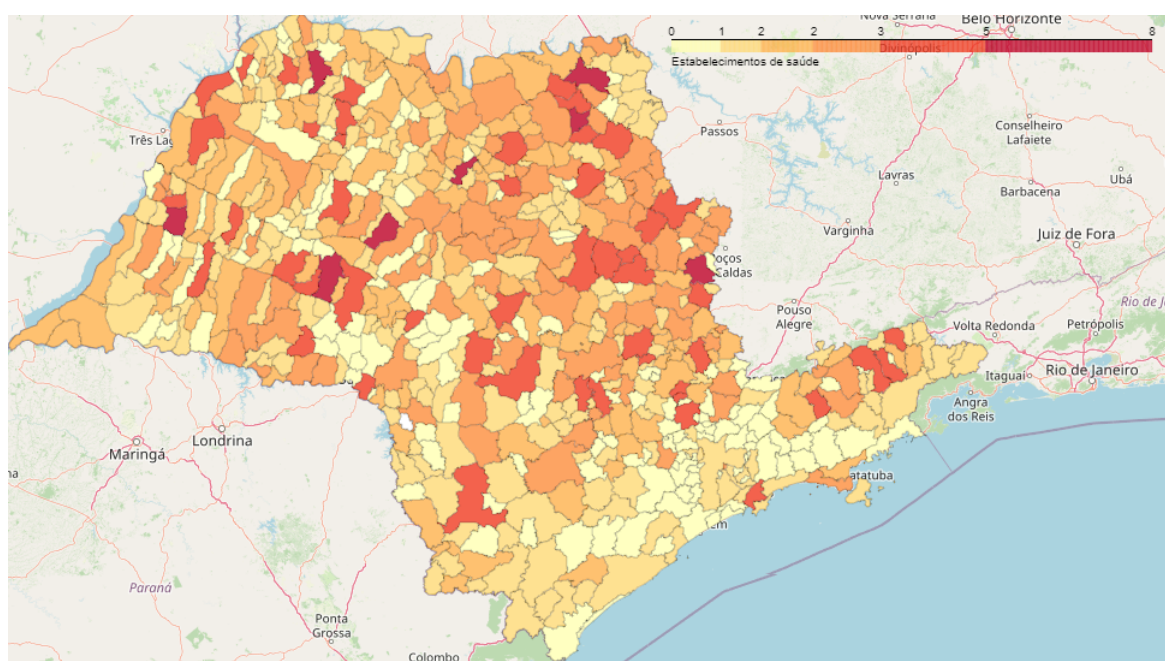
Os três municípios com as maiores taxas de profissionais que não atendem pelo SUS são: Atibaia, Campinas e São Caetano do Sul, com taxas de 3,13, 3,12 e 2,79 respectivamente.

Entre os 10 municípios com as maiores taxas, 8 são caracterizados como cidades médias e 2 são considerados cidades grandes. A média populacional é de 1.544.220 habitantes.

As discrepâncias vão além do quesito populacional, a média IDHM aumenta de 0.764 a 0.811, variando do **alto** para o  **muito alto**. A renda média domiciliar per capita sobe R\$ 481,00, indo de R\$ 843,00 para R\$ 1324,00.

### 4.3 Distribuição de estabelecimentos de saúde

Figura 5 – Quantidade de estabelecimentos a cada 1000 habitantes por município



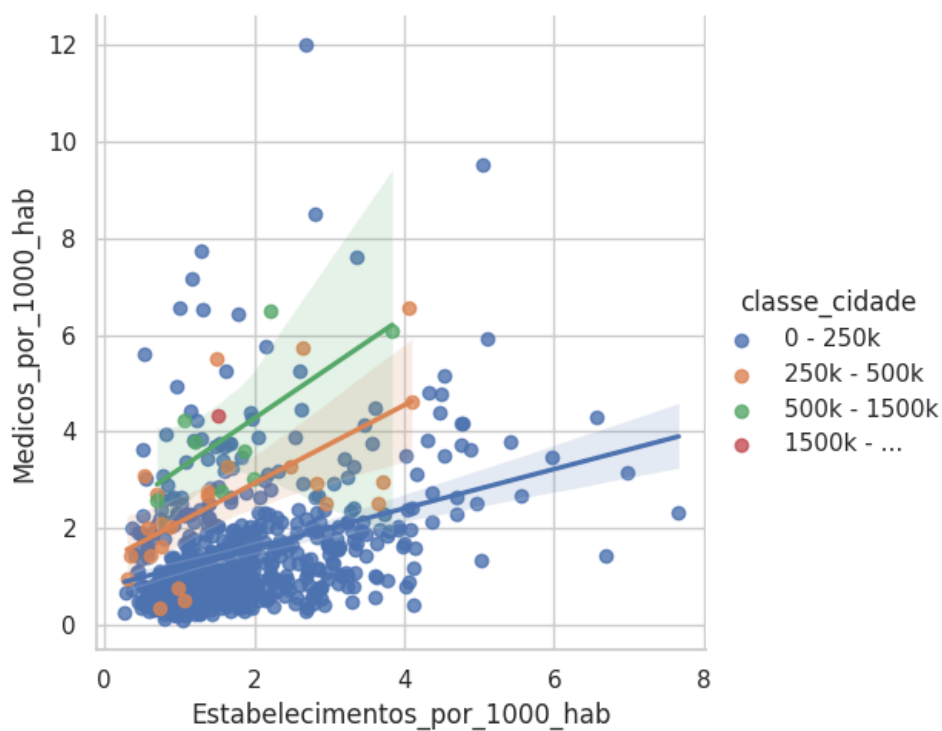
Fonte: Elaborada pelo autor.

A Figura 5 expõe o número de estabelecimentos de saúde atuantes em cada município a cada 1 mil habitantes. A desproporcionalidade com a Figura 1 indica uma eventual desconexão correlacional entre o número de estabelecimentos e capacidade de atendimento, visto que, não foi levado em consideração a magnitude e o tipo da instalação de saúde.

Espacialmente, a disposição abrange começando da região central, avançando para municípios situados ao norte e noroeste do estado. Os municípios de Lins, São João da Boa Vista e Pompéia contam com as maiores taxas, 7,66, 6,97 e 6,68 respectivamente.

Ao analisar os 10 municípios com maior taxa, 8 são considerados cidades pequenas e 2 são considerados cidades médias, com média populacional de 68.857 habitantes e valor médio de IDHM de 0.784, considerado alto. A renda média domiciliar per capita é de R\$ 916,00.

Figura 6 – Médicos x Estabelecimentos



Fonte: Elaborada pelo autor.

O gráfico de dispersão exibido na Figura 6 indica uma correlação positiva entre a quantidade de médicos e o número de estabelecimentos proporcional a população. A partir da legenda, os municípios foram divididos em nível populacional e as linhas de regressão não diferem em níveis notáveis. A correlação pode ser expressa matematicamente ao aplicar a correlação de Pearson, que pode ser visto na tabela 1

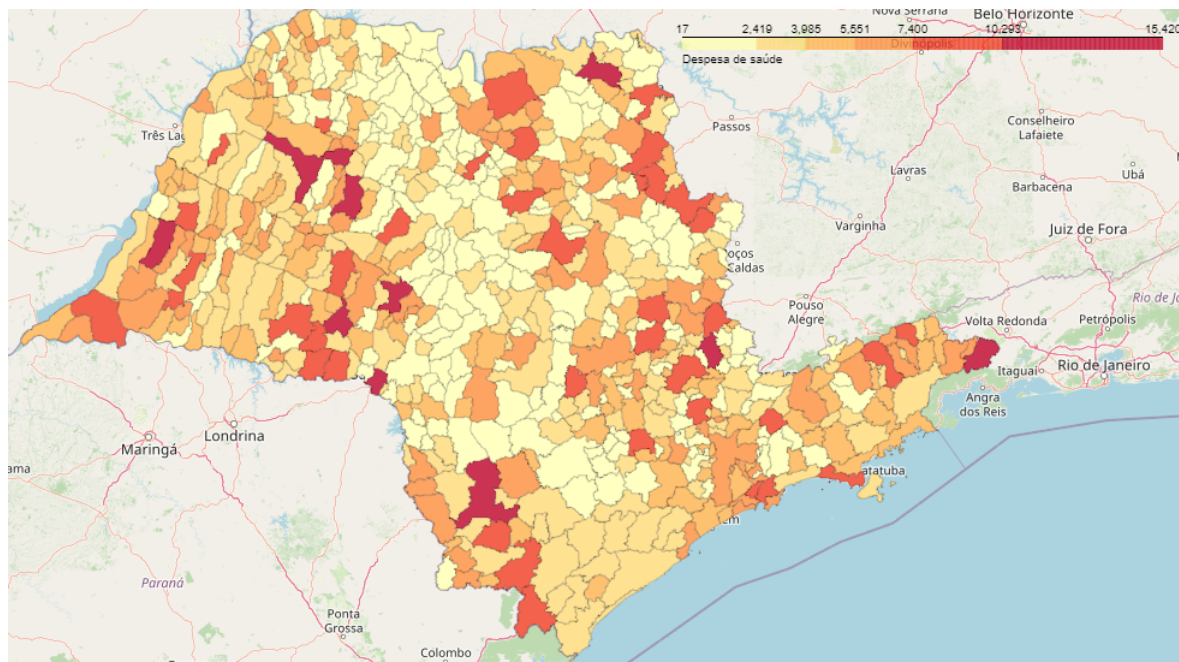
Tabela 1 – Correlação Estabelecimentos x Médicos

|                  | Estabelecimentos | Médicos  |
|------------------|------------------|----------|
| Estabelecimentos | 1.000000         | 0.339725 |
| Médicos          | 0.339725         | 1.000000 |

Fonte: Elaborada pelo autor.

## 4.4 Despesa por município

Figura 7 – Valor despendido a cada 1000 habitantes por município



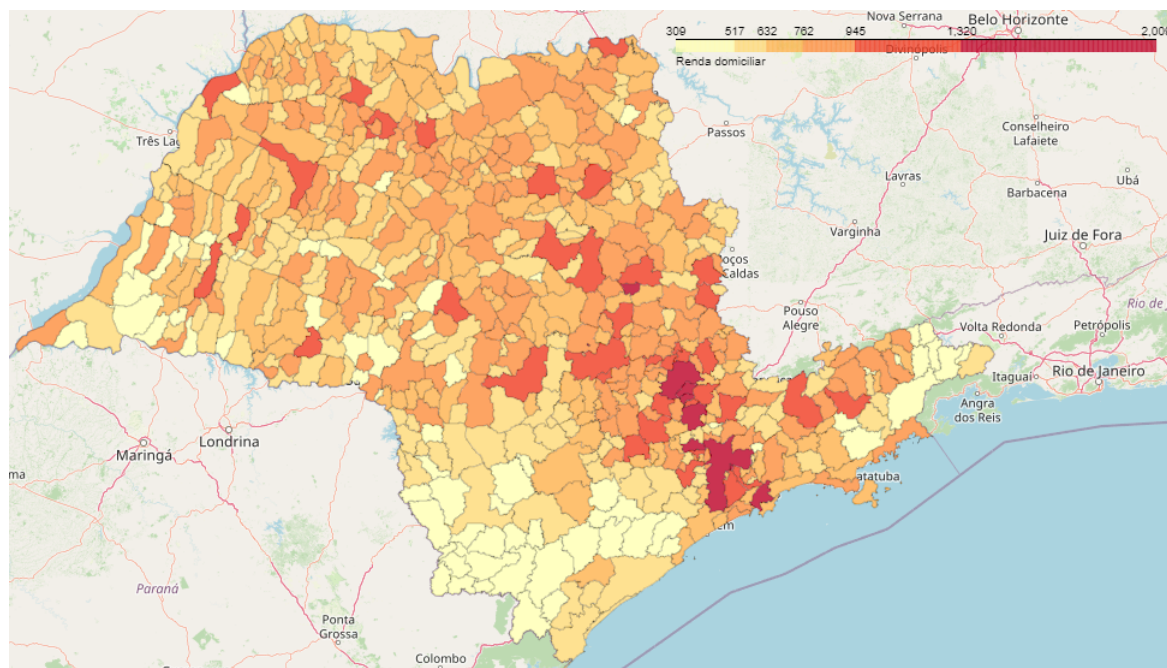
Fonte: Elaborada pelo autor.

A Figura 7 analisa a distribuição proporcional de valor despendido em cada município, isto é, recursos federais utilizados para manter ou melhorar a infraestrutura de saúde da região. É possível perceber que as regiões com maior verba não constituem os maiores conglomerados urbanos do estado, similar ao mapa da Figura 5, que mostra a distribuição de estabelecimentos de saúde a cada 1000 habitantes.

Ao considerar os 10 municípios com maiores valores recebidos, 8 são considerados cidades pequenas e 2 são considerados cidades médias, com média populacional de 62.966 habitantes e valor médio de IDHM de 0.761, considerado alto. A renda média domiciliar per capita é de R\$ 809,00.

## 4.5 Renda domiciliar

Figura 8 – Renda média domiciliar per capita



Fonte: Elaborada pelo autor.

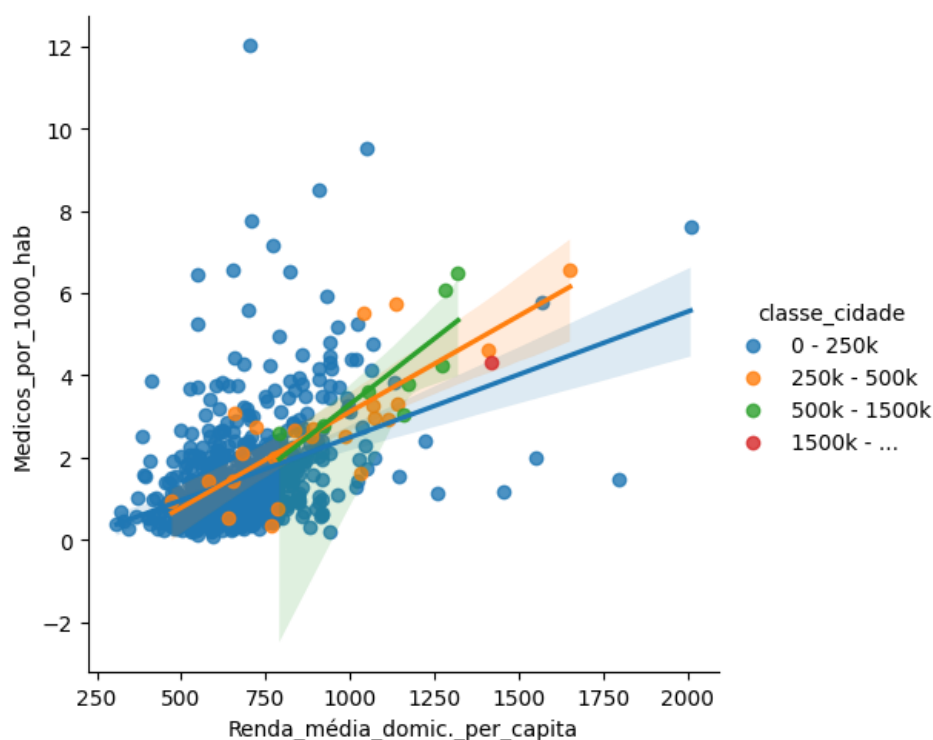
O mapa exposto na Figura 8 mostra a distribuição de renda ao longo do estado. A Renda média domiciliar per capita é calculado a partir da média salarial dos habitantes de um mesmo domicílio, e por vezes, chamado de renda per capita familiar.

A partir do mapa, é possível notar certa relação com as Figuras 1, 2 e 4, que retratam a população, distribuição de médicos e médicos que não atendem pelo SUS, respectivamente, com destaque para a última.

As três cidades com maior renda são: São Caetano do Sul, Santana de Parnaíba e Santos, com valores de R\$ 2.008,98, R\$ 1.798,50 e R\$ 1.651,82 respectivamente. Os dois primeiros municípios integram a região metropolitana de São Paulo e Santos está geograficamente próximo.

Entre os 10 municípios com maior renda, 3 são considerados cidades pequenas, 5 são considerados cidades médias e 2 entram como cidades grandes. A média populacional é de 1.398.108 habitantes e valor médio de IDHM de 0.822, considerado muito alto, o valor médio da renda média domiciliar per capita é de R\$ 1.557,00.

Figura 9 – Médicos x Renda



Fonte: Elaborada pelo autor.

O gráfico da Figura 9 demonstra uma relação positiva entre as duas variáveis, que aumenta conforme avalia-se conglomerados urbanos maiores, entre 250 mil à 1 milhão e meio de habitantes. Matematicamente, pode ser visto de acordo com a tabela 2 um valor próximo de 0.5, maior entre as variáveis apresentadas.

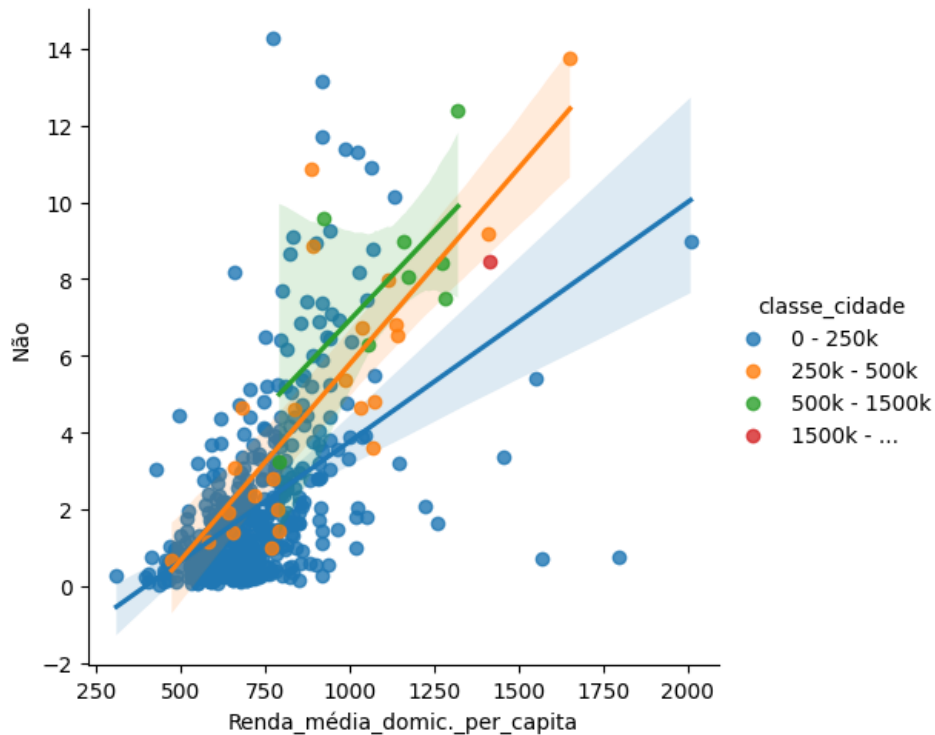
Tabela 2 – Correlação - Médicos x Renda

|         | Renda    | Médicos  |
|---------|----------|----------|
| Renda   | 1.000000 | 0.480376 |
| Médicos | 0.480376 | 1.000000 |

Fonte: Elaborada pelo autor.

Indo além, como mencionado anteriormente, os mapas da Figura 8 e 4, renda média domiciliar per capita e quantidade de médicos a cada 1 mil habitantes por município que não atendem via SUS, respectivamente, exibem semelhança espacial perceptível, o que pode ser expresso graficamente na Figura 10, novamente, com linhas de regressão mais inclinadas conforme aumento populacional.

Figura 10 – Médicos que não atendem via SUS x Renda



Fonte: Elaborada pelo autor.

Os métodos de correlação de Pearson confirmam o realce, nas tabelas 3 e 4 é possível notar a diferença relacional entre a renda média domiciliar per capita e o número de profissionais que atuam ou não via Sistema Único de Saúde.

Tabela 3 – Correlação - Médicos que não atendem via SUS x Renda

|         | Renda    | Médicos  |
|---------|----------|----------|
| Renda   | 1.000000 | 0.601804 |
| Médicos | 0.601804 | 1.000000 |

Fonte: Elaborada pelo autor.

Tabela 4 – Correlação - Médicos que atendem via SUS x Renda

|         | Renda    | Médicos  |
|---------|----------|----------|
| Renda   | 1.000000 | 0.044412 |
| Médicos | 0.044412 | 1.000000 |

Fonte: Elaborada pelo autor.



Nesse cenário, a cidade está exportando pacientes para outras áreas.

O tons de verde indicam um balanço positiva, indo do mais claro até o mais escuro conforme o valor aumenta, e o contrário é verdadeiro, onde os tons de vermelho refletem um balanço de internações negativa.

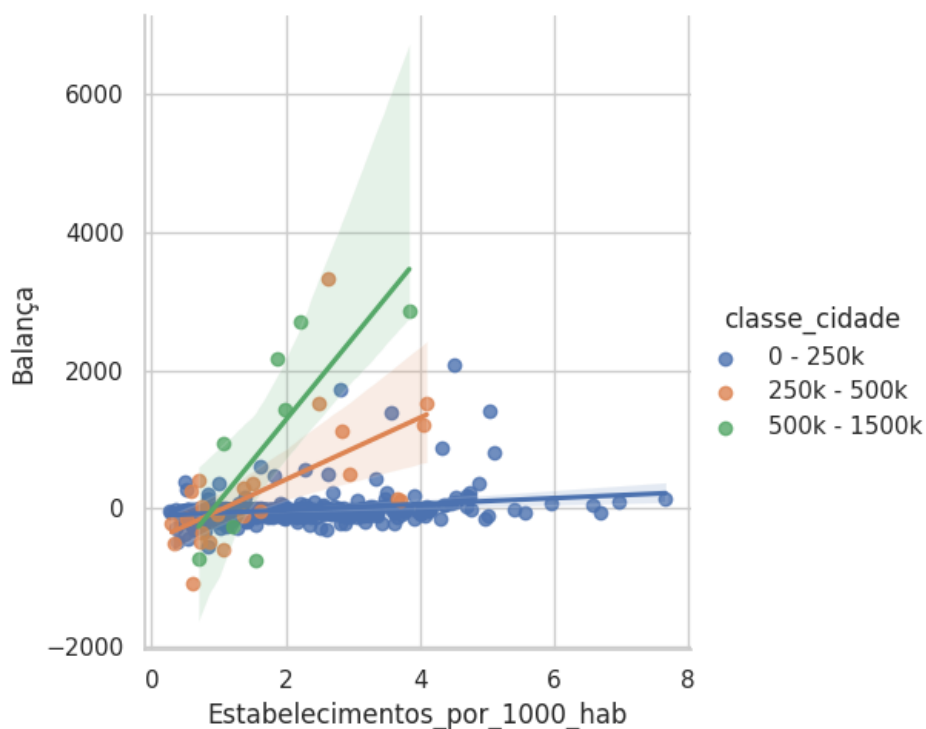
Ao considerar municípios com balanço de internações positiva, em tons esverdeados, é possível estimar que, em média, aproximadamente 27% dos pacientes internados reside em outro município.

As três cidades com maior balanço de internações são: São Paulo, São José do Rio Preto e Ribeirão Preto, com valores de 7.972, 3.321 e 2.861 respectivamente. Geograficamente, não integram a mesma região, porém, constituem-se como as cidades mais populosas dentro de suas respectivas regiões, como pode ser visto na Figura 1.

Entre 10 municípios com maior balanço, 5 são considerados cidades médias e 5 são classificadas como cidades grandes. A média populacional é de 1.635.966 habitantes e valor médio de IDHM de 0.802, considerado muito alto, a renda média domiciliar per capita é de R\$ 1.189,20 e a média do balanço de internações é de 2.729.

Segundo SAUDE (2002), a internação é caracterizada quando um paciente ocupa um leito hospitalar por um período maior ou igual a 24 horas. Os valores contidos no conjunto de dados segue essas diretrizes, logo, para que uma internação aconteça, é necessário infraestrutura adequada para acomodação do paciente. Nesse sentido, cabe analisar a relação entre a distribuição de estabelecimentos de saúde e o balanço de internações dos municípios.

Figura 12 – Balanço x Estabelecimentos / 1000 habitantes



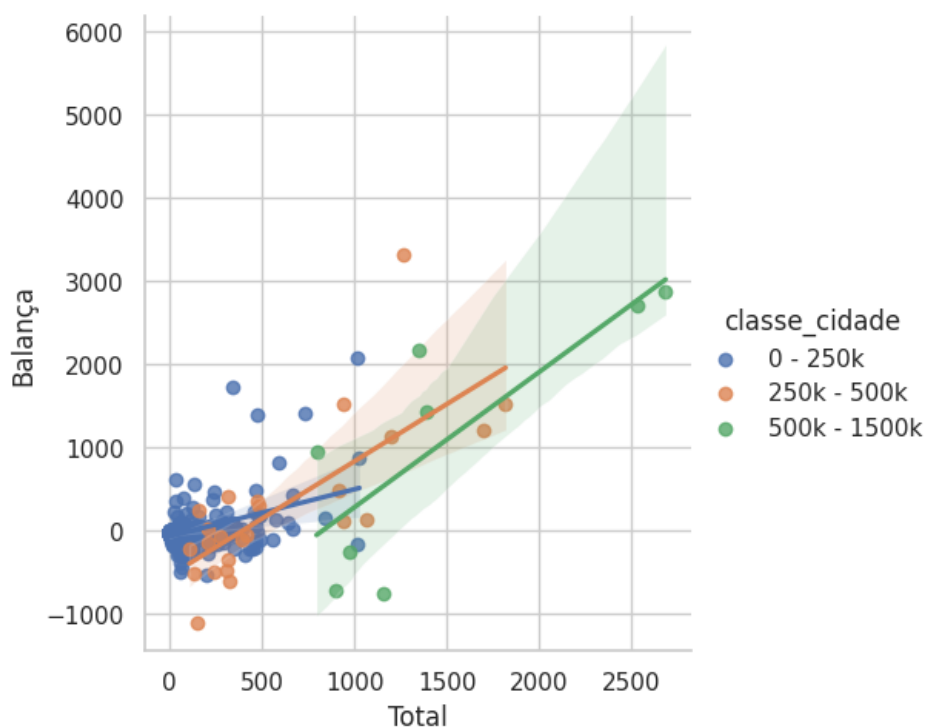
Fonte: Elaborada pelo autor.

O gráfico da Figura 12 relaciona o balanço de interações e o número de estabelecimentos a cada 1000 habitantes.

As cidades com menos habitantes, em azul, mostram fraca correlação entre as variáveis, entretanto, ao isolar os municípios com população maior que 250 mil as linhas de regressão indicam maior relação.

Desprezando a proporcionalidade populacional, ao levar em conta os números absolutos de estabelecimentos, é possível verificar um aumento correlacional geral expressivo em comparação ao gráfico da Figura 12, novamente, ao isolar a análise para municípios com faixa populacional entre 250 mil a 1 milhão e meio, as linhas de regressão indicam maior correlação, como pode ser visto na Figura 13.

Figura 13 – Balanço x Estabelecimentos totais



Fonte: Elaborada pelo autor.

O modelo de Pearson exprime em números essas diferenças, que se mantêm conforme aumento do nível populacional, conforme tabela 5.

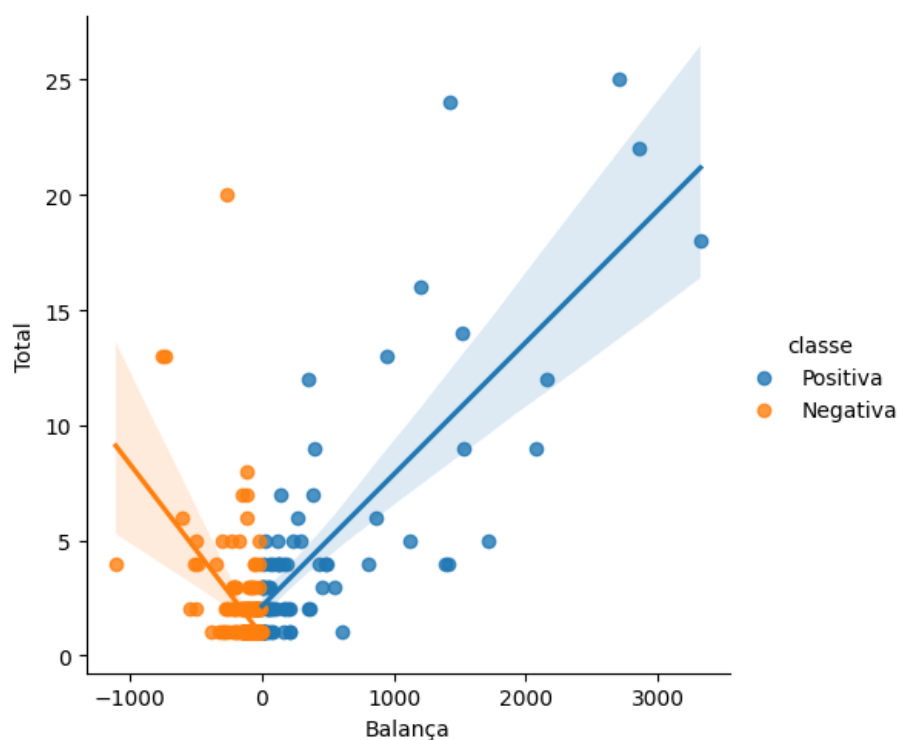
Tabela 5 – Correlação - Balanço x Estabelecimentos

|                            | Balanço  | Estabelecimentos/1000 hab. | Estab. total |
|----------------------------|----------|----------------------------|--------------|
| Balanço                    | 1.000000 | 0.165839                   | 0.827724     |
| Estabelecimentos/1000 hab. | 0.165839 | 1.000000                   | 0.131180     |
| Estab. total               | 0.827724 | 0.131180                   | 1.000000     |

Fonte: Elaborada pelo autor.

Também é possível filtrar os estabelecimentos, incluindo apenas hospitais e pronto-socorros gerais, considerando que nem todo estabelecimento tem o objetivo de ou possui infraestrutura para internações de pacientes. O gráfico na Figura 14 relaciona a quantidade de estabelecimentos com maior infraestrutura ao balanço de internações do município.

Figura 14 – Balanço x Hospitais/Pronto-Socorros



Fonte: Elaborada pelo autor.

A partir do gráfico é possível perceber um vínculo semelhante ao observado no gráfico da Figura 13, que pode ser validado com a tabela 6.

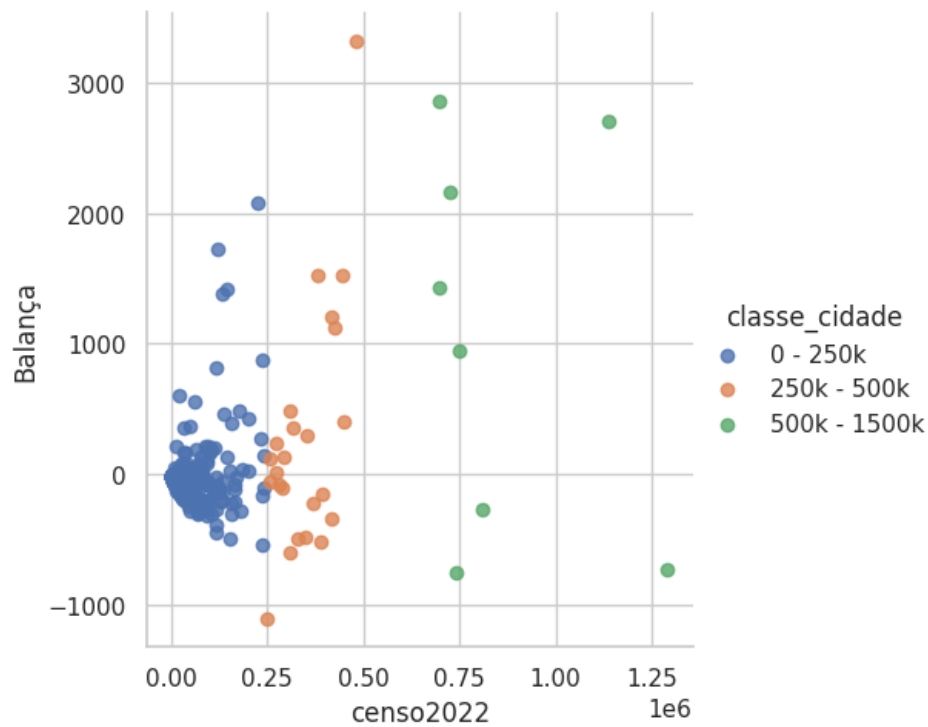
Tabela 6 – Correlação - Balanço x Hospitais/Pronto-Socorros

|           | Balanço  | Hospitais |
|-----------|----------|-----------|
| Balanço   | 1.000000 | 0.798783  |
| Hospitais | 0.798783 | 1.000000  |

Fonte: Elaborada pelo autor.

Como os gráficos das Figuras 12 e 13 indicam influência populacional, a Figura 15 permite avaliar o reflexo da quantidade de habitantes de cada município. Ao avaliar o gráfico junto com a Tabela 7 é possível identificar a correlação e entender os resultados apresentados nas duas Figuras anteriores ao variar o nível populacional.

Figura 15 – Balanço x Habitantes



Fonte: Elaborada pelo autor.

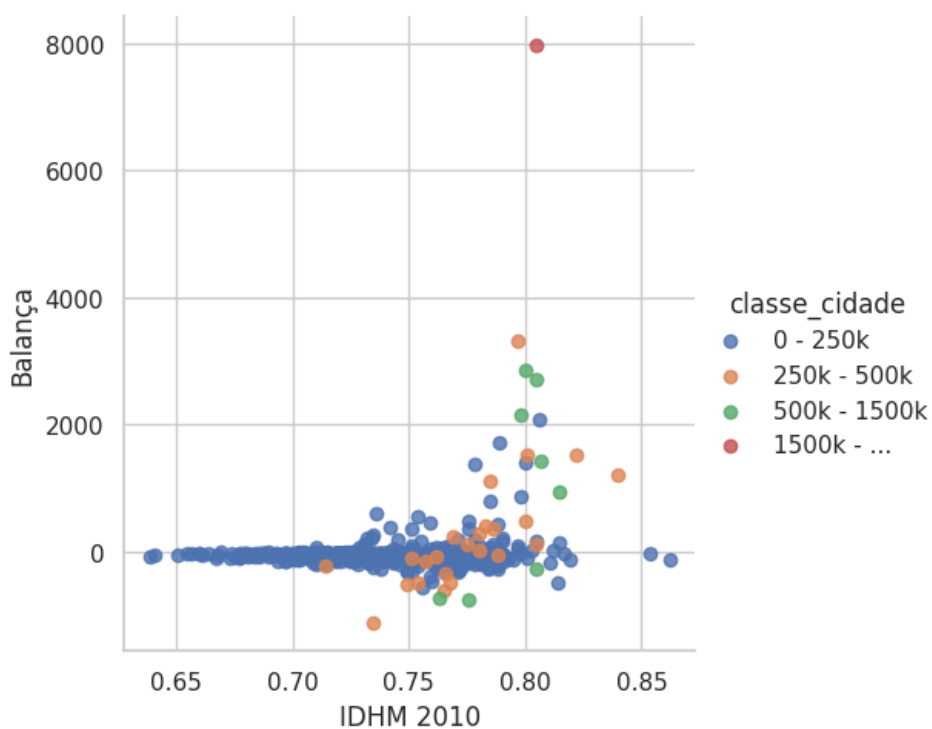
Tabela 7 – Correlação - Balanço x População

|           | Balanço  | População |
|-----------|----------|-----------|
| Balanço   | 1.000000 | 0.756754  |
| População | 0.756754 | 1.000000  |

Fonte: Elaborada pelo autor.

Os quatro níveis do Índice de Desenvolvimento Humano (IDH) vão do **baixo**, até  **muito alto**, entre os municípios do estado de São Paulo, varia-se do nível **médio** até o **muito alto**. Conforme o gráfico da Figura 16, é perceptível a concentração de municípios com população até 250 mil com valores de balanço de interações próximo a 0, indicando independência ou pouca dependência entre as variáveis, o que pode ser confirmado ao aplicar o método de Pearson na tabela 8.

Figura 16 – Balanço x IDHM



Fonte: Elaborada pelo autor.

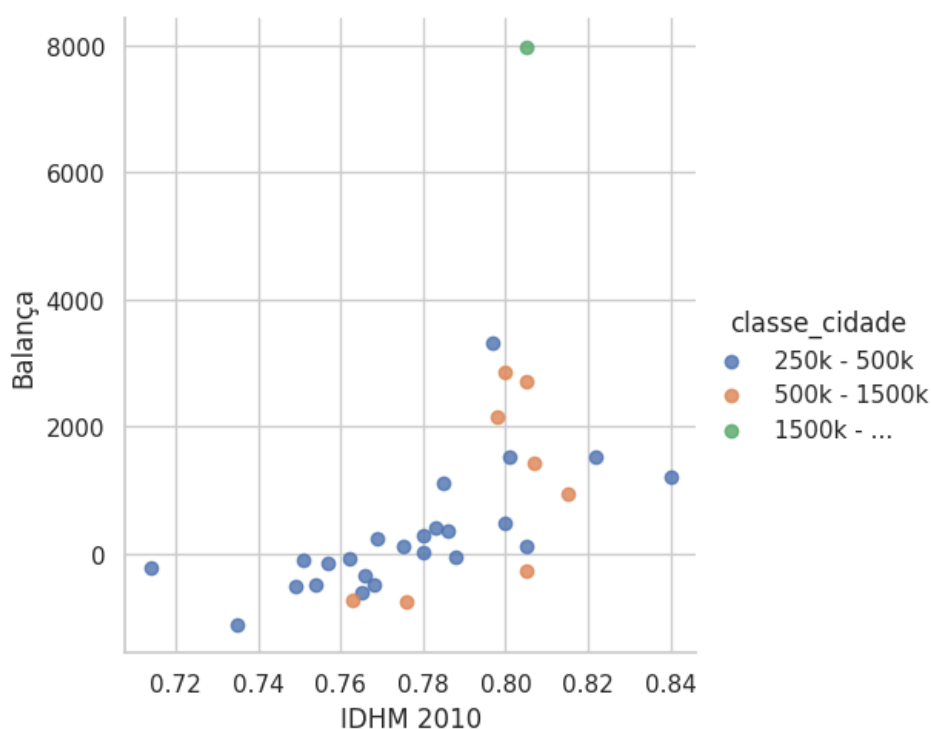
Tabela 8 – Correlação - Balanço x IDH municipal

|         | Balanço  | IDHM     |
|---------|----------|----------|
| Balanço | 1.000000 | 0.242424 |
| IDHM    | 0.242424 | 1.000000 |

Fonte: Elaborada pelo autor.

Ao filtrar os municípios, excluindo cidades com menos de 250 mil habitantes, nota-se um aumento na relação entre o IDH e o balanço de internações, como pode ser visto na Figura 17 e na tabela 9. Apontando, outra vez, que o nível populacional do município em questão corrobora para um aumento no balanço de internações.

Figura 17 – Balanço x IDHM - Pop. &gt; 250k



Fonte: Elaborada pelo autor.

Tabela 9 – Correlação - Balanço x IDHM - Pop &gt; 250k

|         | Balanço  | IDHM     |
|---------|----------|----------|
| Balanço | 1.000000 | 0.525593 |
| IDHM    | 0.525593 | 1.000000 |

Fonte: Elaborada pelo autor.

## 4.7 Considerações

É essencial destacar algumas observações-chave. As análises realizadas, incluindo as análises geoespaciais, gráficas e correlacionais, proporcionaram uma visão abrangente do cenário de saúde nos municípios do estado de São Paulo. Embora tenham explorado aspectos diferentes, estão interligadas por um objetivo comum: compreender os desafios da saúde buscando futuras melhorias nas políticas públicas.

Em relação à presença de relações, foi identificado correlações significativas entre variáveis importantes, como renda per capita, número de médicos, deslocamento de pacientes e população. Essas conexões ressaltam a complexidade dos fatores que influenciam a distribuição de médicos e o acesso aos serviços de saúde. No entanto, vale ressaltar que, embora as análises tenham revelado correlações, elas não implicam necessariamente em causalidade, e é necessário considerar outras variáveis não analisadas que podem contribuir para esses resultados.

As análises geoespaciais revelaram que a má distribuição de médicos é mais acentuada em regiões específicas do estado, principalmente nas regiões fora de conglomerados urbanos, como o oeste e sul do estado. As análises gráficas mostraram que está correlacionada com fatores como renda média domiciliar per capita e população e a divergência no número de internações pode vir de variáveis como número de estabelecimentos de saúde, população e IDH municipal. As análises correlacionais matemáticas confirmam essas correlações.

Em relação às cidades ou regiões destacadas, observou-se que algumas se destacam de forma positiva, indicando uma melhor distribuição de recursos médicos e menor deslocamento de pacientes, enquanto outras se destacam negativamente, sinalizando a necessidade de intervenção para melhorar o acesso à assistência médica.

As cidades ou regiões que se destacam de forma positiva ou negativa na maioria dos casos são aquelas que apresentam características específicas que influenciam a distribuição de médicos. As cidades ou regiões com maior renda per capita, IDH municipal e densamente povoadas tendem a ter uma melhor distribuição de médicos e infraestrutura médica.

## 5 Conclusão

No presente trabalho foi realizada uma exploração das questões de saúde pública no estado de São Paulo, abordando a desigualdade na distribuição de médicos e o alto deslocamento de pacientes para internações. Ao longo deste estudo, foi empregado uma variedade de técnicas de análise, incluindo análises geoespaciais, análises gráficas e análises correlacionais matemáticas pelo método de Pearson, com o objetivo de identificar padrões e relações cruciais.

As análises geoespaciais revelaram uma imagem abrangente da distribuição de médicos em todo o estado de São Paulo. Os mapas forneceram uma representação visual clara das áreas com excesso e escassez de profissionais de saúde. Também permitiram identificar áreas onde os pacientes enfrentam deslocamentos substanciais para receber atendimento médico, o que, por sua vez, impacta negativamente a acessibilidade aos serviços de saúde.

A utilização de gráficos de dispersão nos possibilitou explorar relações entre variáveis-chave, como renda per capita, número de médicos e deslocamento de pacientes. Essa abordagem resultou na identificação de correlações preliminares entre esses fatores. Além disso, a análise de correlação de Pearson forneceu medidas quantitativas da força e direção das relações entre as variáveis, conferindo maior robustez aos resultados.

Em último ponto, a análise de dados da saúde apresentada neste trabalho não apenas revelou as complexidades do sistema de saúde em São Paulo, mas também apontou direções para a melhoria das políticas públicas. O caminho adiante envolve uma colaboração contínua entre pesquisadores, governo e profissionais de saúde, com o objetivo comum de garantir um sistema de saúde mais equitativo e acessível para todos os cidadãos do estado de São Paulo.

### 5.1 Trabalhos Futuros

É fundamental reconhecer que, apesar dos resultados obtidos, ainda há espaço para o aprimoramento das técnicas utilizadas. Outras metodologias estatísticas e de análise de dados podem ser exploradas para uma compreensão mais profunda das questões de saúde pública no estado. Ademais, a inclusão de dados adicionais, pode enriquecer as análises e a compreensão das raízes dos desafios enfrentados.

Os desafios da desigualdade na distribuição de médicos e do deslocamento de pacientes para internações não possuem soluções únicas. Este estudo fornece uma base sólida para futuras pesquisas e iniciativas de políticas públicas. A exploração de estratégias como incentivos à alocação de médicos em áreas carentes e o aprimoramento da infraestrutura de saúde pode contribuir para mitigar os problemas identificados.

## Referências

CNES. 2023. Disponível em: <<[https://wiki.saude.gov.br/cnes/index.php/P%C3%A1gina\\_principal#Principais\\_T.C3.B3picos\\_do\\_CNES](https://wiki.saude.gov.br/cnes/index.php/P%C3%A1gina_principal#Principais_T.C3.B3picos_do_CNES)>>. Acesso em 17 de Outubro de 2023.

DATASUS. s.d. Disponível em: <<<https://datasus.saude.gov.br/sobre-o-datasus/>>>. Acesso em 20 de Outubro de 2023.

OCDE. *Who we are*. s.d. Disponível em: <<<https://www.oecd.org/about/>>>. Acesso em 18 de Outubro de 2023.

PEREIRA, S. D. *Conceitos e Definições da Saúde e Epidemiologia usados na Vigilância Sanitária*. 2007. Disponível em: <<[https://cvs.saude.sp.gov.br/pdf/epid\\_visa.pdf](https://cvs.saude.sp.gov.br/pdf/epid_visa.pdf)>>. Acesso em 05 de Outubro de 2023.

SAUDE, M. da. *Padronização da Nomenclatura do Censo Hospitalar*. 2002. Disponível em: <<[http://bvsms.saude.gov.br/bvs/publicacoes/padronizacao\\_censo.pdf](http://bvsms.saude.gov.br/bvs/publicacoes/padronizacao_censo.pdf)>>. Acesso em 12 de Outubro de 2023.

SCHEFFER, M.; GUILLOUX, A. G. A.; MIOTTO, B. A.; ALMEIDA, C. de J. Demografia médica no brasil 2023. 2023. Disponível em: <[https://amb.org.br/wp-content/uploads/2023/02/DemografiaMedica2023\\_8fev-1.pdf](https://amb.org.br/wp-content/uploads/2023/02/DemografiaMedica2023_8fev-1.pdf)> . Acesso em: 11 de abril de 2023.

SES-MG. *SUS*. 2015. <<https://www.saude.mg.gov.br/sus>>. Acesso em: 07 de abril de 2023.

SUS. s.d. Disponível em: <<<https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/s/sus>>>. Acesso em 09 de Outubro de 2023.