

**UNIVERSIDADE DE SÃO PAULO**

Instituto de Ciências Matemáticas e de Computação

**Agrupamento de séries temporais das taxas de mortalidade por armas de fogo dos municípios do Estado de São Paulo: uma aplicação do algoritmo K-Means**

**Tiago Silva Birkholz Duarte**

Monografia - MBA em Ciência de Dados (CEMEAI)



SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: \_\_\_\_\_

**Tiago Silva Birkholz Duarte**

## **Agrupamento de séries temporais das taxas de mortalidade por armas de fogo dos municípios do Estado de São Paulo: uma aplicação do algoritmo K-Means**

Monografia apresentada ao Centro de Ciências Matemáticas Aplicadas à Indústria do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Ciências de Dados.

Área de concentração: Ciências de Dados

Orientador: Prof. Dr. Afonso Paiva Neto

**Versão original**

**São Carlos**

**2026**

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi  
e Seção Técnica de Informática, ICMC/USP,  
com os dados inseridos pelo(a) autor(a)

D812a Duarte, Tiago Silva Birkholz  
Agrupamento de séries temporais das taxas de mortalidade por armas de fogo dos municípios do Estado de São Paulo: uma aplicação do algoritmo K-Means / Tiago Silva Birkholz Duarte; orientador Afonso Paiva Neto. -- São Carlos, 2026.  
52 p.

Trabalho de conclusão de curso (MBA em Ciência de Dados) -- Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2026.

1. Armas de fogo. 2. Taxas de mortalidade. 3. Séries temporais. 4. K-Means. 5. Algoritmo de agrupamento. I. Paiva Neto, Afonso, orient. II. Título.

**Tiago Silva Birkholz Duarte**

**Clustering of Firearm Mortality Rate Time Series in Cities  
of São Paulo State: An Application of the K-Means  
Algorithm**

Monograph presented to the Centro de Ciências Matemáticas Aplicadas à Indústria do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, as part of the requirements for obtaining the title of Specialist in Data Science.

Concentration area: Data Science

Advisor: Prof. Dr. Afonso Paiva Neto

**Original version**

**São Carlos**

**2026**



## **AGRADECIMENTOS**

Concluir esse MBA em Ciências de Dados e finalizar o presente Trabalho de Conclusão de Curso não foram tarefas simples. Recebi diversas formas de apoio para conseguir chegar até este ponto. Portanto, registro aqui meus sinceros agradecimentos.

Em primeiro lugar, à Secretaria da Fazenda e Planejamento do Estado de São Paulo pelo incentivo aos estudos como forma de qualificação do setor público e pelo financiamento desse curso de especialização.

Também sou grato ao professor Afonso Paiva Neto, que, desde a primeira reunião de orientação, foi acolhedor e paciente com um aluno que não é da área. Com uma comunicação acessível, uma organização pragmática e um rigor acadêmico ímpar, foi essencial para que esse TCC chegasse a termo. Muito obrigado!

Aos professores e monitores do MBA em Ciências de Dados, com os quais aprendi técnicas e modelos diversos para resolver uma ampla variedade de problemas. Em especial, as aulas didáticas e instigantes dos professores Cibele Russo e Francisco Rodrigues me marcaram sobremaneira. Certamente, essas aulas serão consultadas com frequência no futuro.

Aos colegas de curso, que transformaram um grupo de Whatsapp em um verdadeiro grupo de estudos e plantão de dúvidas 24 horas por dia. Tivemos uma interação muito valiosa e educativa.

À minha família e aos amigos, pela compreensão e suporte nesse período de muitas ausências.

E à Dalila, pela companhia, ternura e rosnadas diárias.



## RESUMO

DUARTE, T.S.B. **Agrupamento de séries temporais das taxas de mortalidade por armas de fogo dos municípios do Estado de São Paulo: uma aplicação do algoritmo K-Means.** 2026. 52 p. Monografia (MBA em Ciências de Dados) - Centro de Ciências Matemáticas Aplicadas à Indústria, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2026.

Este trabalho buscou verificar o comportamento das séries temporais de mortalidade por armas de fogo no Estado de São Paulo, um óbito evitável por meio de políticas públicas específicas. Seu objetivo foi elaborar taxas anuais de mortalidade por essa causa nos municípios de São Paulo e criar agrupamentos desses municípios segundo similaridades em suas séries, utilizando técnicas de aprendizado de máquina. Um referencial teórico sobre o tema e uma revisão da bibliografia sobre tendências dessa mortalidade no Brasil foram apresentados. Os dados anuais de mortalidade por armas de fogo e de população municipal foram adquiridos via Datasus. Todos os bancos de dados passaram por pré-processamento, com filtragem por município e causa básica de óbito, bem como recodificação de diversas variáveis. A periodicidade estudada foi dos anos 2000 a 2023. Nesse período, houve 123.725 óbitos por armas de fogo entre residentes do Estado de São Paulo. Ocorreram cerca de 15 óbitos de homens para cada óbito de mulher; 56,7% das mortes atingiram jovens de 15 a 29 anos de idade; e 86,1% foram homicídios. Visando diminuir o efeito de small data, optou-se por comparar apenas municípios de maior porte: as cem cidades paulistas mais populosas em 2023. Foram ajustadas taxas de mortalidade brutas (TMB) e taxas de mortalidade padronizadas por idade e sexo (TMP) para cada ano e município. TMB e TMP apresentaram diferenças discretas, ambas com média de 11,8 óbitos e mediana de 8,5 óbitos por 100 mil habitantes para os municípios-ano. A TMP foi a base para a aplicação de Dynamic Time Warping (DTW), medida de similaridade adequada para séries temporais. Com a métrica de distância, foi realizado um agrupamento hierárquico com o método de ligação por média e apresentado seu dendrograma. A distância por DTW também serviu para agrupar as séries temporais municipais com o algoritmo de clusterização K-Means. A escolha do número de clusters levou em conta o método do cotovelo (inércia) e o escore de silhueta, no qual se buscou o máximo local. O melhor K revelou outliers conjunturais, e o modelo de agrupamento final teve seis clusters e dois municípios não agrupados. O modelo apontou quais cidades possuem comportamentos temporais semelhantes para a mortalidade por armas de fogo. Taxas médias de cada cluster foram calculadas para resumir o comportamento dos grupos. Essas séries médias foram suavizadas com Média Móvel Exponencialmente Ponderada e passaram por testes de estacionariedade. Houve uma tendência geral de declínio nas taxas de mortalidade por armas de fogo no período, mas com diferenças relevantes entre os clusters. Nos grupos 2, 3 e 6 (que contam com municípios da Região Metropolitana de São Paulo e alguns

grandes centros regionais do interior), as quedas foram claras, com mais força e com mais persistência do que nos grupos 1 e 4. Já no grupo 5, não houve clareza sobre tendência. As duas séries outliers correspondem a municípios do Vale do Paraíba. Conclui-se que a redução na mortalidade por armas de fogo não foi homogênea entre as cidades de São Paulo, e as evidências permitem que gestores públicos possam proceder com medidas focalizadas para alguns municípios e com políticas melhoradas de prevenção à violência letal armada.

**Palavras-chave:** Armas de fogo. Taxas de mortalidade. Séries temporais. K-Means. Algoritmo de agrupamento.

## ABSTRACT

DUARTE, T.S.B. **Clustering of Firearm Mortality Rate Time Series in Cities of São Paulo State: An Application of the K-Means Algorithm.** 2026. 52 p. Monograph (MBA in Data Sciences) - Centro de Ciências Matemáticas Aplicadas à Indústria, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2026.

This study aimed to examine the behavior of firearm mortality time series in the State of São Paulo, an avoidable cause of death through specific policies. Its objective was to estimate annual mortality rates from this cause in São Paulo municipalities and to create clusters of these municipalities according to similarities in their time series, using machine learning techniques. A theoretical framework on the topic and a review of the literature on trends in firearm mortality in Brazil were presented. Annual data on firearm-related deaths and municipal population data were obtained from Datasus. All datasets underwent preprocessing, including filtering by municipality and underlying cause of death, as well as the recoding of several variables. The study period covered the years from 2000 to 2023. During this period, there were 123,725 firearm-related deaths among residents of the State of São Paulo. Approximately 15 male deaths occurred for each female death; 56.7% of deaths affected young people aged 15 to 29 years; and 86.1% were homicides. In order to reduce the effects of small data, the analysis was restricted to larger municipalities: the 100 most populous cities in São Paulo in 2023. Crude mortality rates (CMR) and age- and sex-standardized mortality rates (SMR) were calculated for each year and municipality. CMR and SMR showed minor differences, both presenting a mean of 11.8 deaths and a median of 8.5 deaths per 100,000 inhabitants for municipality-years. SMR served as the basis for the application of Dynamic Time Warping (DTW), a similarity measure suitable for time series. Using this distance metric, hierarchical clustering was performed with the average linkage method, and its dendrogram was presented. DTW distance was also used to cluster municipal time series through the K-Means clustering algorithm. The choice of the number of clusters considered the elbow method (inertia) and the silhouette score, seeking a local maximum. The optimal K revealed contextual outliers, and the final clustering model consisted of six clusters and two ungrouped municipalities. The model identified cities with similar temporal patterns of firearm mortality. Mean rates for each cluster were calculated to summarize group behavior. These mean series were smoothed using an Exponentially Weighted Moving Average and subjected to stationarity tests. A general downward trend in firearm mortality rates was observed over the period, with relevant differences among clusters. In clusters 2, 3, and 6 – which include municipalities from the São Paulo Metropolitan Region and major regional centers in the state's interior – the declines were clearer, stronger, and more persistent than in clusters 1 and 4. In cluster 5, no clear trend was identified. The two outlier series correspond to municipalities in

the Vale do Paraíba. The results indicate that the reduction in firearm mortality was not homogeneous across São Paulo cities, and the evidence allows public managers to adopt targeted measures for specific municipalities and improved policies for the prevention of lethal armed violence.

**Keywords:** Firearms. Mortality rates. Time series. K-Means. Clustering algorithm.

## LISTA DE FIGURAS

Figura 1 – Distribuição das populações dos cem municípios mais populosos de São Paulo em 2023 (até percentil 99). . . . .	34
Figura 2 – Distribuição das taxas anuais médias de mortalidade bruta (TMB) e padronizada por idade e sexo (TMP) dos cem municípios mais populosos de São Paulo. . . . .	35
Figura 3 – Distribuição das taxas de mortalidade padronizadas por idade e sexo (TMP) dos dez municípios mais populosos de São Paulo, 2000-2023. . . . .	36
Figura 4 – Taxas de mortalidade padronizadas por idade e sexo (TMP) dos dez municípios mais populosos de São Paulo, 2000-2023. . . . .	37
Figura 5 – Heatmap da matriz de distâncias DTW dos dez municípios mais populosos de São Paulo. . . . .	38
Figura 6 – Dendrograma dos cem municípios mais populosos de São Paulo (método de ligação por média). . . . .	39
Figura 7 – Método do cotovelo (inércia). . . . .	40
Figura 8 – Escore de silhueta. . . . .	40
Figura 9 – Taxas de mortalidade padronizadas por idade e sexo (TMP), 2000-2023. Curvas individuais de cada cluster com distância DTW. . . . .	41
Figura 10 – Taxas médias de mortalidade padronizadas por idade e sexo (TMP média), 2000-2023. Curvas médias de cada cluster com distância DTW. . . . .	42
Figura 11 – Tendência suavizada (MMEP) das taxas médias de mortalidade padronizadas por idade e sexo (TMP média), 2000-2023. Curvas médias de cada cluster com distância DTW. . . . .	42
Figura 12 – Clusters de mortalidade por armas de fogo. Cem municípios mais populosos do Estado de São Paulo. . . . .	44



## LISTA DE TABELAS

Tabela 1 – Trabalhos identificados na revisão bibliográfica de tendências de séries temporais das taxas de mortalidade violenta. . . . .	28
Tabela 2 – Estatísticas descritivas dos municípios-ano, referentes aos cem municípios mais populosos de São Paulo no período 2000-2023. . . . .	35
Tabela 3 – Municípios agrupados segundo clusters obtidos por K-Means com DTW.	43



## LISTA DE ABREVIATURAS E SIGLAS

AAPC	Average Annual Percent Change
APC	Annual Percent Change
ADF	Dickey-Fuller Aumentado
CID-10	Classificação Estatística Internacional de Doenças e Problemas Relacionados com a Saúde – 10 <sup>a</sup> Revisão
Datasus	Departamento de Informação e Informática do SUS
DBA	DTW Barycenter Averaging
DTW	Dynamic Time Warping
IBGE	Instituto Brasileiro de Geografia e Estatística
IC 95%	Intervalo de Confiança de 95%
ICMC-USP	Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo
IIQ	Intervalo Interquartil
MMEP	Média Móvel Exponencialmente Ponderada
RMSP	Região Metropolitana de São Paulo
SFP	Secretaria da Fazenda e Planejamento
SIM	Sistema de Informações sobre Mortalidade
TCC	Trabalho de Conclusão de Curso
TMB	Taxa de Mortalidade Bruta
TMP	Taxa de Mortalidade Padronizada por Idade e Sexo
WPGMA	Weighted Pair Group Method with Arithmetic Mean



## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> . . . . .	<b>21</b>
<b>1.1</b>	<b>Motivação</b> . . . . .	<b>22</b>
<b>1.2</b>	<b>Objetivos</b> . . . . .	<b>22</b>
1.2.1	Objetivos específicos . . . . .	22
<b>2</b>	<b>REFERENCIAL TEÓRICO E REVISÃO BIBLIOGRÁFICA</b> . . . . .	<b>23</b>
<b>2.1</b>	<b>Referencial teórico</b> . . . . .	<b>23</b>
<b>2.2</b>	<b>Revisão bibliográfica</b> . . . . .	<b>25</b>
<b>3</b>	<b>METODOLOGIA</b> . . . . .	<b>29</b>
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b> . . . . .	<b>33</b>
<b>4.1</b>	<b>Análise exploratória dos dados</b> . . . . .	<b>33</b>
<b>4.2</b>	<b>Agrupamento de séries temporais</b> . . . . .	<b>37</b>
<b>4.3</b>	<b>Discussão sobre o agrupamento de séries temporais</b> . . . . .	<b>45</b>
<b>5</b>	<b>CONCLUSÃO</b> . . . . .	<b>49</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>51</b>



## 1 INTRODUÇÃO

A mortalidade por armas de fogo é um relevante problema público no mundo e, em particular, no Brasil. Globalmente, 40% dos homicídios ocorrem por armas de fogo; na América do Sul, essa proporção é muito maior, de 70%. O Brasil é um país com uma alta taxa de homicídios. Trata-se da segunda maior da América do Sul, e grande parte desses óbitos decorre do uso de armas de fogo (United Nations Office on Drugs and Crime, 2023). As armas de fogo também são responsáveis por suicídios, fenômeno que ocorre em menor proporção no Brasil quando comparado com o volume de homicídios. No entanto, os suicídios merecem igual atenção por conta de sua tendência recente de crescimento no Brasil (Rodrigues *et al.*, 2019).

Como as armas de fogo são um meio importante para o óbito decorrente da agressão perpetrada por terceiros (homicídio) ou para o óbito autoinfligido (suicídio), além de ser também responsável por óbitos acidentais e aqueles cuja intenção não pôde ser determinada pelas autoridades competentes, propõe-se, neste Trabalho de Conclusão de Curso (TCC), realizar um recorte das mortes por armas de fogo e avaliar seu comportamento no Estado de São Paulo.

Tal proposta apresenta uma abordagem pouco usual sobre esse tipo de mortalidade evitável, pois a maioria dos trabalhos sobre séries temporais de mortalidade trata ou de homicídios, ou de suicídios, no Brasil ou em seus entes federados. Ambos são fenômenos de violência distintos, com causas e contextos particulares. Contudo, na ausência da disponibilidade de uma arma de fogo, ambos poderiam ter um desfecho não letal. Portanto, a análise agregada, incluindo homicídios, suicídios, acidentes e mortes com intenção não determinada, permitiria demonstrar um quadro mais abrangente da violência decorrente do uso de armas de fogo e se enquadra na agenda de políticas públicas de controle e restrição desse tipo de instrumento.

## 1.1 Motivação

Uma investigação acerca da mortalidade por armas de fogo, em particular do comportamento das séries históricas desse óbito, é pertinente, afinal trata-se de um fenômeno evitável por meio de políticas públicas específicas, que abrevia a vida de um contingente expressivo de pessoas no país. Considerando que a participação do autor no Curso de Especialização EAD MBA em Ciências de Dados do Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (ICMC-USP) e a realização deste TCC decorrem de financiamento da Secretaria da Fazenda e Planejamento (SFP) do Estado de São Paulo para seus servidores públicos, o projeto focalizou o problema estudado no Estado de São Paulo, com análises performadas para os municípios paulistas, de acordo com interesses dos gestores da área de avaliação de políticas públicas da SFP e do autor.

## 1.2 Objetivos

Neste trabalho, propõe-se elaborar as séries temporais das taxas de mortalidade por armas de fogo dos municípios do Estado de São Paulo e criar agrupamentos desses municípios segundo similaridades em suas taxas, utilizando técnicas de aprendizado de máquina.

### 1.2.1 Objetivos específicos

- Elaborar as taxas de mortalidade por armas de fogo para os municípios paulistas a partir dos dados de mortalidade e de população;
- Realizar análise exploratória dos dados brutos e das taxas de mortalidade elaboradas;
- Testar modelos de agrupamento das séries temporais das taxas de mortalidade;
- Aplicar o melhor modelo de agrupamento para identificar municípios paulistas com comportamentos semelhantes nas séries temporais das taxas de mortalidade.

## 2 REFERENCIAL TEÓRICO E REVISÃO BIBLIOGRÁFICA

Neste capítulo, serão apresentados conceitos e um referencial teórico para melhor compreensão do tema do presente trabalho, bem como uma sucinta revisão da bibliografia sobre as tendências da mortalidade por armas de fogo no Brasil.

### 2.1 Referencial teórico

Usualmente, chamamos de mortalidade por causas externas (ou mortalidade por causas não naturais) aqueles óbitos cuja causa básica é um acidente ou uma violência. Violências podem ser aquelas autoinfligidas (que resultam no óbito por suicídio) ou infligidas por terceiros (que resultam no óbito por homicídio).

Causa básica da morte é um conceito fundamental para diversas áreas de políticas públicas e campos científicos, tais como, a saúde pública, a vigilância epidemiológica, a justiça, a segurança pública e a sociologia, e para as estatísticas de mortalidade em geral. Trata-se do elemento que identifica o motivo original do óbito, sendo definido internacionalmente como:

“A doença ou lesão que iniciou a cadeia de acontecimentos patológicos que conduziram diretamente à morte, ou as circunstâncias do acidente ou violência que produziram a lesão fatal.” (Organização Mundial da Saúde, 2022).

O Brasil conta com um sistema de registro de óbitos, chamado Sistema de Informações sobre Mortalidade (SIM), existente desde 1975 e sob responsabilidade do Ministério da Saúde. Este é alimentado com dados provenientes da Declaração de Óbito, uma ficha nacional padronizada que, dentre outras informações, tem o registro da causa básica (Brasil. Ministério da Saúde, 2024). Esse registro é realizado segundo a codificação da Classificação Estatística Internacional de Doenças e Problemas Relacionados com a Saúde – 10<sup>a</sup> Revisão (CID-10).

Para este trabalho, como estamos compilando causas externas de óbito distintos, com a finalidade de agrupar todos aqueles decorrentes de armas de fogo, utilizaremos os códigos da CID-10 referentes às causas básicas de morte de diferentes tipos, mas cujo meio identificado foi sempre o uso de arma de fogo, ou seja, as seguintes mortes por causa externa com uso de arma de fogo e respectivos códigos da CID-10: homicídios (X93-X95), suicídios (X72-X74), acidentes (W32-W34), óbitos cuja intenção não foi determinada (Y22-Y24) e intervenções legais (Y35.0).

O evento de interesse em estudos desse tipo não é o número de óbitos, mas a taxa de mortalidade, métrica que utiliza como denominador a população sob risco. Segundo o Dicionário de Epidemiologia (Porta *et al.*, 2014), a taxa de mortalidade é:

“Uma estimativa da proporção de uma população que morre durante um determinado período. O numerador é o número de pessoas que morreram durante o período; o denominador é o número de pessoas na população, geralmente estimado como a população do meio do ano. A taxa de mortalidade em uma população é geralmente calculada pela seguinte fórmula:

$$\left( \frac{\text{Número de mortes durante um período específico}}{\text{Número de pessoas em risco de morrer durante o período}} \right) \times 10^n$$

Essa taxa é uma estimativa da taxa de mortalidade por pessoa-tempo, ou seja, a taxa de mortalidade por 10<sup>n</sup> pessoas-ano. Se a taxa for baixa, também é uma boa estimativa da taxa de mortalidade cumulativa. Essa taxa também é chamada de taxa bruta de mortalidade.” (Tradução nossa).

Portanto, além dos dados sobre óbitos, adquiridos por meio do SIM, utilizaremos estimativas populacionais anuais dos municípios, de modo a elaborar as taxas de mortalidade bruta e padronizada por idade e sexo. Esses dados de estimativas populacionais anuais dos municípios são disponibilizados no site do Ministério da Saúde, assim como os dados de óbitos do SIM.

## 2.2 Revisão bibliográfica

Como mencionado na introdução, homicídios e suicídios são fenômenos de violência distintos. Não possuem as mesmas causas, não ocorrem sob os mesmos contextos. Mas a presença ou não de uma arma de fogo naqueles contextos, entretanto, pode ser a diferença entre um desfecho letal ou não letal do ato violento. Portanto, essa revisão bibliográfica apresenta estudos sobre esses tipos de violência no cenário brasileiro e aqueles que tratam especificamente de análises de tendência das taxas de ambos os tipos de óbitos, destacando aqueles em que o objetivo (ou, pelo menos, uma das análises secundárias) explorou a morte decorrente do uso de arma de fogo.

No Brasil nos anos 1990, foi constatado um crescimento relevante nas taxas de mortalidade por homicídios, com significativa contribuição do uso de armas de fogo nesse crescimento (Peres; Santos, 2005). Entre 2001 e 2015, foram contabilizados 778.450 homicídios no país (quantidade de óbitos que supera conflitos de independência ou guerra civil), dos quais 70,8% ocorreram com a presença de armas de fogo; dos 138.086 suicídios no mesmo período, 12,2% tiveram a participação de armas de fogo (Jorge, 2020).

O crescimento no homicídio brasileiro também foi descrito por Murray e col. (2013), em uma série de 1980 a 2010, onde se nota o crescimento estável de suas taxas nos anos 1990 e início dos anos 2000. Esse estudo destaca a proporção de homicídios por armas de fogo. A partir de 1990, o uso dessas armas passa a ser o predominante em relação a todos os outros meios, elevando as taxas até o pico da série, em 2003. A partir de 2004, a série se estabiliza até 2010, mudança que coincide com a implementação de relevante política pública brasileira no contexto do controle de armas (Murray; Cerqueira; Kahn, 2013).

O controle de armas de fogo como meio para a diminuição dos homicídios é tema controverso na literatura internacional. Ainda assim, por conta da elevada violência e da tendência de crescimento dos homicídios no país, foi criada a Lei nº 10.826/2003, conhecida como o Estatuto do Desarmamento. Dentre outras medidas, a nova legislação criou uma Campanha do Desarmamento em 2004-2005, que foi seguida por outras duas campanhas em anos posteriores, as quais juntas retiraram mais de 660 mil armas de circulação no Brasil (Jorge, 2020).

Para Santiago e col. (2021), entre 2000 e 2019, as tendências nas taxas de homicídio (aqui incluindo óbitos por intervenções legais e eventos de intenção não determinada) apresentaram padrões distintos entre as regiões brasileiras. Houve, em geral, um crescimento nas taxas das regiões Norte e Nordeste, bem como queda na taxa da região Sudeste, a mais populosa do país. No Brasil de 2007 a 2017, houve um crescimento significativo dos homicídios (APC = 2,0%,  $p < 0,001$ ), seguido de queda (APC = -15,5%,  $p < 0,001$ ). Nos 20 anos analisados, constatou-se também aumento significativo nas mortes por intervenção legal (AAPC = 19,3%, IC 95% [11,5; 27,7]), isto é, decorrentes da ação de policiais ou

outros agentes legais (Santiago; Nunes; Macena, 2021). Nesse estudo, não houve distinção nos meios perpetrados para o homicídio, portanto há óbitos que não foram causados por armas de fogo. Contudo, a maioria dos homicídios no Brasil ocorre com armas de fogo, como mencionado anteriormente, assim como as mortes por intervenção legal.

O estudo de Malta e col. (2020), que correlaciona diversas taxas de mortalidade por armas de fogo com taxas de devolução voluntária de armas de fogo, apresentou uma conclusão interessante que apoia a linha da política de desarmamento: após o início da coleta voluntária de armas de fogo, estados brasileiros que coletaram mais armas, reduziram suas taxas de morte violenta por esses instrumentos. No período entre 2000 e 2017, os estados com maior taxa de coleta (mediana = 7,5 armas por 100 mil habitantes, intervalo interquartil [3,5; 12,5]) apresentaram reduções nas taxas de mortes totais por armas de fogo (APC = -1,5%) e de homicídios por armas de fogo (APC = -1,6%). Enquanto os estados com menor coleta (mediana = 2,0, IIQ [0,5; 4,5]), sofreram um crescimento nas taxas de mortes totais por armas de fogo (APC = 3,3%) e de homicídios por armas de fogo (APC = 4,1%), apesar de haver desaceleração no período imediatamente posterior ao Estatuto do Desarmamento (Malta *et al.*, 2020).

Um trabalho que tratou dos óbitos por arma de fogo, mas de maneira mais restrita – tanto do ponto de vista do objetivo (tratava-se apenas do homicídio de mulheres de 15 a 49 anos de idade), quanto de método (optou-se apenas por um código da CID-10 referente ao homicídio por armas de fogo, X93) – foi o de Soares e col. (2024). Não foram encontradas tendência para as taxas no Brasil e nas regiões no período estudado (2010 a 2019), exceto uma queda na região Centro-Oeste entre 2010 e 2014 (APC = -41,9%, IC 95% [-64,7; -4,3]), bem como um crescimento na região Norte entre 2013 e 2019 (APC = 23,5%, IC 95% [0,4; 51,9]) (Soares *et al.*, 2024).

Em estudo semelhante no objetivo, mas mais robusto na metodologia, Meira e col. (2024) analisam a tendência temporal dos homicídios femininos (a partir de 10 anos de idade) perpetrados por arma de fogo nos estados nordestinos entre os anos de 2000 e 2019. Em todos os estados, constataram-se tendências ascendentes nas taxas de homicídios femininos com uso de arma de fogo, com exceção de Pernambuco (tendência descendente) e Alagoas (tendência estacionária). Na região Nordeste como um todo a tendência foi ascendente com AAPC = 3,58% (IC 95% [2,11; 5,06]) (Meira *et al.*, 2024).

Historicamente, os suicídios ocorrem em proporção muito menor no país do que os homicídios. Ainda assim, o uso de armas de fogo é um meio relevante na consecução desse ato violento. Quanto às taxas de suicídio por armas de fogo no Brasil, houve declínio para homens e mulheres entre 2000 e 2017, segundo McDonald e col. (2021). Há diferenças bastante nítidas na magnitude do fenômeno do suicídio entre homens e mulheres, assim as análises ocorreram com essa estratificação. Para mulheres, a queda nas taxas ocorreu entre 2000 e 2011 (APC = -7,3%, IC 95% [-8,4; -6,2]), permanecendo estacionária em

seguida. Para homens, a estabilidade esteve no primeiro período, de 2000 a 2002, depois houve quedas em duas intensidades: de 2002 a 2010 (APC = -6,4%, IC 95% [-7,7; -5,0]) e de 2010 a 2017 (APC = -2,5%, IC 95% [-4,1; -0,1]). Os autores consideram as quedas nas taxas específicas por armas de fogo consistentes com a implementação do Estatuto do Desarmamento (McDonald *et al.*, 2021). Importante destacar que, nesse período, as taxas de suicídio em geral cresceram no Brasil (Rodrigues *et al.*, 2019; McDonald *et al.*, 2021).

Na capital paulista, a mesma tendência de queda nos suicídios por armas de fogo ocorreu entre 1996 e 2018 para homens e mulheres. Entre os homens – e considerando que, nesse grupo, as armas são o segundo meio mais comum para a consecução do suicídio –, houve uma queda nas taxas entre 1996 e 2001 (APC = -11,3%, IC 95% [-15,7; -6,6]), crescimento até 2006 (APC = 4,9%, IC 95% [0,4; 9,6]) e nova queda até o fim da série em 2018 (APC = -5,6%, IC 95% [-7,4; -3,7]). Nas mulheres, a tendência de declínio foi constante em todo o período da série temporal (APC = -6,6%, IC 95% [-9,0; -4,1]) (Duarte, 2022).

As análises brasileiras de séries temporais nos campos da saúde pública, sociologia e economia normalmente utilizam-se de análises meramente descritivas ou de modelos estatísticos clássicos de séries temporais para ajustar as tendências da mortalidade por homicídio ou suicídio. A Tabela 1 resume os trabalhos identificados nesta revisão (descartados estudos meramente descritivos), apontando tanto o método utilizado na análise, quanto o resultado sobre tendência das taxas de mortalidade.

Apesar de não terem sido identificados trabalhos sobre tendências das taxas de mortalidade utilizando algoritmos de aprendizado de máquina, outros estudos do campo já fazem uso de metodologias com esses algoritmos para atingir objetivos distintos. No artigo de Alves e col. (2018) sobre predição do crime, a variável dependente (target) utilizada – sua única definição de crime – foi o número de homicídios em nível municipal no Brasil. Um regressor de floresta aleatória criou predições, a partir de indicadores urbanos, para os homicídios das cidades com uma acurácia acima de 97% da variância explicada (Alves; Ribeiro; Rodrigues, 2018).

Com proposta semelhante para a cidade de Belém (Pará), Ribeiro e col. (2021) utilizaram um classificador de floresta aleatória para prever se em um dado bairro da cidade no mês subsequente haveria ou não um homicídio, considerando como variáveis para o treinamento do algoritmo as quantidades de outros crimes que ocorrem em cada mês em cada bairro (Ribeiro *et al.*, 2021).

Para o contexto norte-americano, há trabalhos com o uso de algoritmos mais rebuscados no tema específico da violência por armas de fogo, por exemplo: a criação de um índice de vulnerabilidade de violência por armas de fogo com base em incidentes de tiro em comunidades utilizando XGBoost (Polcari *et al.*, 2023); estimativas em tempo quase real para os homicídios por armas de fogo semanais e anuais nos Estados Unidos

elaboradas com o uso de modelos de ensemble (LASSO) (Swedo *et al.*, 2023); e a utilização de ElasticNet para identificar quais políticas públicas melhor predizem as taxas de suicídio por armas de fogo em nível estadual (Goldstein; Wilson, 2024).

Por fim, há propostas de modelagem preditiva com técnicas de aprendizado de máquina para séries temporais das taxas de mortalidade, tais como o uso de aprendizado profundo (Perla *et al.*, 2021) e de análise dinâmica de componentes principais funcionais com estratégias de previsão multi-step-ahead (DirRec) (Beyaztas; Shang, 2022).

No presente trabalho, iremos proceder com uma forma diferenciada de análise de séries temporais, utilizando algoritmos de agrupamento para verificar similaridades entre as curvas das taxas de mortalidade por armas de fogo dos municípios de São Paulo.

Tabela 1 – Trabalhos identificados na revisão bibliográfica de tendências de séries temporais das taxas de mortalidade violenta.

Referência	Fenômeno estudado	Localidade	Período	Método empregado	Resultado
Santiago e col. (2021)	Homicídios em geral, incluindo intervenções legais	Brasil e regiões	2000-2019	Joinpoint Regression Program v. 4.9.0.0	Crescimento até 2017 seguido de queda no Brasil
Malta e col. (2020)	Mortes por armas de fogo entre estados com maior e menor coleta de armas	Estados brasileiros	2000-2017	Regressão linear	Decréscimo para estados com mais coleta de armas, crescimento para os outros
Soares e col. (2024)	Homicídios de mulheres (15 a 49 anos) por armas de fogo	Brasil e regiões	2010-2019	Joinpoint Regression Program v. 4.6.0.0	Sem tendência para Brasil
Meira e col. (2024)	Homicídios de mulheres ( $\geq 10$ anos) por armas de fogo	Estados do Nordeste	2000-2019	Regressão binomial negativa	Crescimento na região Nordeste e maioria dos estados
McDonald e col. (2021)	Suicídio por armas de fogo entre homens e mulheres	Brasil	2000-2017	Joinpoint Regression Program v. 4.8.0.1	Decréscimo para ambos os sexos
Duarte (2022)	Suicídio por armas de fogo entre homens e mulheres	Município de São Paulo	1996-2018	Regressão de Poisson	Decréscimo para ambos os sexos

### 3 METODOLOGIA

Primeiramente, foram apresentados alguns conceitos necessários para o entendimento das taxas de mortalidade e foi realizado um levantamento bibliográfico sobre o problema da mortalidade por armas de fogo no Brasil, sua magnitude e tendências, apontando as técnicas utilizadas nos trabalhos e os tipos de ocorrência do óbito, tais como homicídios e suicídios.

Os dados mais recentes de mortalidade por armas de fogo para o Estado de São Paulo foram adquiridos por meio do Sistema de Informação sobre Mortalidade (SIM) do Departamento de Informação e Informática do SUS (Datasus). Para tanto, foi desenvolvido um pipeline para localização no FTP do Datasus, identificação, download e filtragem dos bancos de mortalidade de cada ano. Como os bancos referem-se ao Brasil, com identificação por municípios, filtrou-se pela variável “município de residência”, visando compilar somente os dados de ocorrências de residentes no Estado de São Paulo, isto é, casos com o código de valor 35 nos dois primeiros dígitos da variável “município de residência”.

Também foi realizado o filtro pelas causas básicas de óbito de interesse, a saber, óbitos por armas de fogo. Essas são identificadas pelos códigos da CID-10 na variável causa básica da morte. Os códigos associados foram aqueles referentes a: disparo de armas de fogo acidental (W32, W33, W34), lesão autoprovocada intencionalmente (suicídio) por disparo de armas de fogo (X72, X73, X74), agressão (homicídio) por meio de disparo de armas de fogo (X93, X94, X95), disparo de armas de fogo com intenção não determinada (Y22, Y23, Y24) e intervenção legal envolvendo o uso de armas de fogo (Y350).

Os dados de mortalidade passaram por um pré-processamento, que transformou e recodificou as variáveis: causa básica da morte (agrupando os tipos de causa), ano do óbito, idade (do formato original do Datasus, que possui três dígitos, sendo o primeiro referente à unidade da idade, para a idade em anos e a idade em faixas etárias), sexo (recodificando valores faltantes) e município de residência (recodificando municípios não identificados, casos em que o óbito foi registrado apenas como de residente do Estado de São Paulo). Ainda que sigam o padrão do código do Instituto Brasileiro de Geografia e Estatística (IBGE), os códigos dos municípios nas bases do SIM apresentam diferenças entre os anos: há casos com os códigos de seis dígitos e outros com sete dígitos. Todos os dados de identificação do município de residência foram padronizados com seis dígitos.

Visando a elaboração de taxas anuais, são necessários dados populacionais dos municípios paulistas. De modo semelhante ao realizado nos bancos relativos aos óbitos, os bancos anuais de população foram localizados no FTP do Datasus, identificados, baixados e filtrados a cada ano. Houve pré-processamento das variáveis ano, idade (criação de faixas

etárias iguais àquelas criadas para os óbitos) e município de residência (recodificando os códigos de municípios para seis dígitos).

Muitos municípios podem apresentar, pela própria natureza do fenômeno ou por conta de sua baixa população, poucos casos de óbitos anuais por armas de fogo. Isso implica em problemas de small data, isto é, quando as estimativas das taxas se tornam instáveis e pouco confiáveis, sujeitas a grandes variações relativas ao ocorrer uma pequena variação absoluta no numerador. Portanto, optamos por comparar apenas municípios de maior porte do Estado de São Paulo. A análise apenas com cidades mais populosas visa diminuir o efeito da eventual flutuação aleatória decorrente de small data. Desse modo, identificamos os municípios com maiores populações no ano mais recente da série temporal (em 2023) e definimos os cem mais populosos para compor este trabalho. A menor cidade dentre as cem possui população superior a 75 mil habitantes.

A escolha metodológica pelo período de série temporal deste TCC (dos anos 2000 a 2023) decorre de: 1) haver o mesmo modo de codificação de doenças e problemas de saúde, isto é, utilização da CID-10, codificação que se inicia em 1996 e ainda é utilizada em 2023, ano dos últimos dados de mortalidade disponíveis; 2) disponibilidade de estimativas populacionais com as variáveis necessárias para a elaboração de taxas e que segue o mesmo padrão de estimação em todo o período: o primeiro banco desse tipo disponível refere-se ao ano 2000.

As estratificações da população segundo faixa etária e sexo foram utilizadas para a criação de taxas de mortalidade padronizadas por idade e sexo (TMP). O uso dessas taxas tem como objetivo controlar diferenças na composição demográfica entre os municípios analisados, isto é, permite eliminar vieses decorrentes das distintas estruturas etárias e de sexo das populações municipais, tornando as taxas comparáveis entre si, independentemente de tempo ou de localidade. Essas taxas foram estimadas pelo método direto, utilizando como população padrão o Censo Demográfico do Brasil de 2010. A decisão pelo Censo 2010 se justifica por: 1) usar dados censitários é melhor do que usar dados populacionais estimados; 2) o ano de 2010 é intermediário, considerando o período deste estudo, o que apoia a redução de viés temporal; 3) o uso do censo nacional é procedimento comum em padronização de taxas de mortalidade, permitindo comparações com outros estudos epidemiológicos.

Para a elaboração das taxas pelo método direto, inicialmente, calcularam-se as taxas específicas de mortes por armas de fogo por faixa etária e sexo para cada município. Em seguida, essas taxas foram ponderadas pelos pesos correspondentes da população padrão e somadas, resultando em uma taxa ajustada única por município, expressa em 100.000 habitantes.

As taxas de mortalidade anuais (brutas e padronizadas) foram ajustadas para cada município no período 2000 a 2023. Esses dados são a base para a aplicação da técnica

---

de Dynamic Time Warping (DTW), visando identificar séries cujas características mais importantes sejam seu formato geral. DTW é uma medida de similaridade adequada para séries temporais, que define uma distância entre elas. Para tanto, o DTW requer que: cada ponto de uma série temporal seja combinado com pelo menos um ponto de outra série temporal; o primeiro e último índice de cada série deve ser comparado com os equivalentes das outras séries temporais; o mapeamento dos pontos deve ser de tal forma que o tempo avance e não retroceda (Nielsen, 2021).

A similaridade encontrada pelo DTW serviu para o agrupamento das séries temporais. Após a aplicação dessas medidas, utilizamos K-Means, um popular algoritmo de aprendizado não supervisionado para clusterização. Há uma versão adaptada de K-Means para séries temporais, que utiliza a métrica DTW para esse algoritmo no módulo clustering da biblioteca tslearn. O uso da distância euclidiana para agrupar com K-Means, no caso de séries temporais, não é recomendado (Tavenard *et al.*, 2020).

O algoritmo de K-Means realiza iterações sucessivas de atribuição das séries e atualização dos centroides: primeiro, cada série temporal é associada ao centroide do cluster cuja distância DTW é mínima; depois, cada centroide é recalculado a partir das séries que lhe foram atribuídas, por meio de um método (DTW Barycenter Averaging, DBA) que gera uma média temporal levando em conta o alinhamento não linear entre as séries. Após atingir a convergência, o algoritmo terá agrupado cada série em um dos K clusters. A definição do tamanho do hiperparâmetro K decorreu de análises do dendrograma, do método do cotovelo (inércia) e do score de silhueta com o DTW, no qual se buscou o máximo local.

Os agrupamentos definidos serviram para apontar quais municípios paulistas possuem comportamentos semelhantes nas séries temporais das taxas de mortalidade por armas de fogo. Foram elaboradas taxas médias de cada cluster, de modo a resumir o comportamento de cada grupo. Estas foram suavizadas por meio de Média Móvel Exponencialmente Ponderada (MMEP), com uma janela de cinco anos, para apresentar, com menos ruído, a tendência das curvas de cada cluster. As taxas médias de cada cluster também passaram por testes de estacionariedade (Teste de Dickey-Fuller Aumentado - ADF). Ao final, uma discussão sobre os resultados da clusterização foi desenvolvida.



## 4 RESULTADOS E DISCUSSÃO

Neste capítulo, será apresentada a análise exploratória dos dados de mortalidade por armas de fogo, de população e das taxas de mortalidade brutas (TMB) e padronizadas por idade e sexo (TMP) dos municípios paulistas. Em seguida, será realizada a análise de agrupamento das séries temporais das taxas de mortalidade por armas de fogo e seus resultados serão discutidos.

### 4.1 Análise exploratória dos dados

A compilação dos dados de mortalidade por armas de fogo no Estado de São Paulo revelou 123.725 óbitos ocasionados por esse meio entre 2000 e 2023. Destes, 2.907 não tiveram a especificação do município de residência, havendo apenas o registro do óbito como referente a um morador do Estado de São Paulo. Os outros 120.818 óbitos ocorreram entre residentes de 612 municípios, dos 645 municípios paulistas existentes. Logo, 94,9% dos municípios paulistas tiveram algum óbito por arma de fogo nesse período.

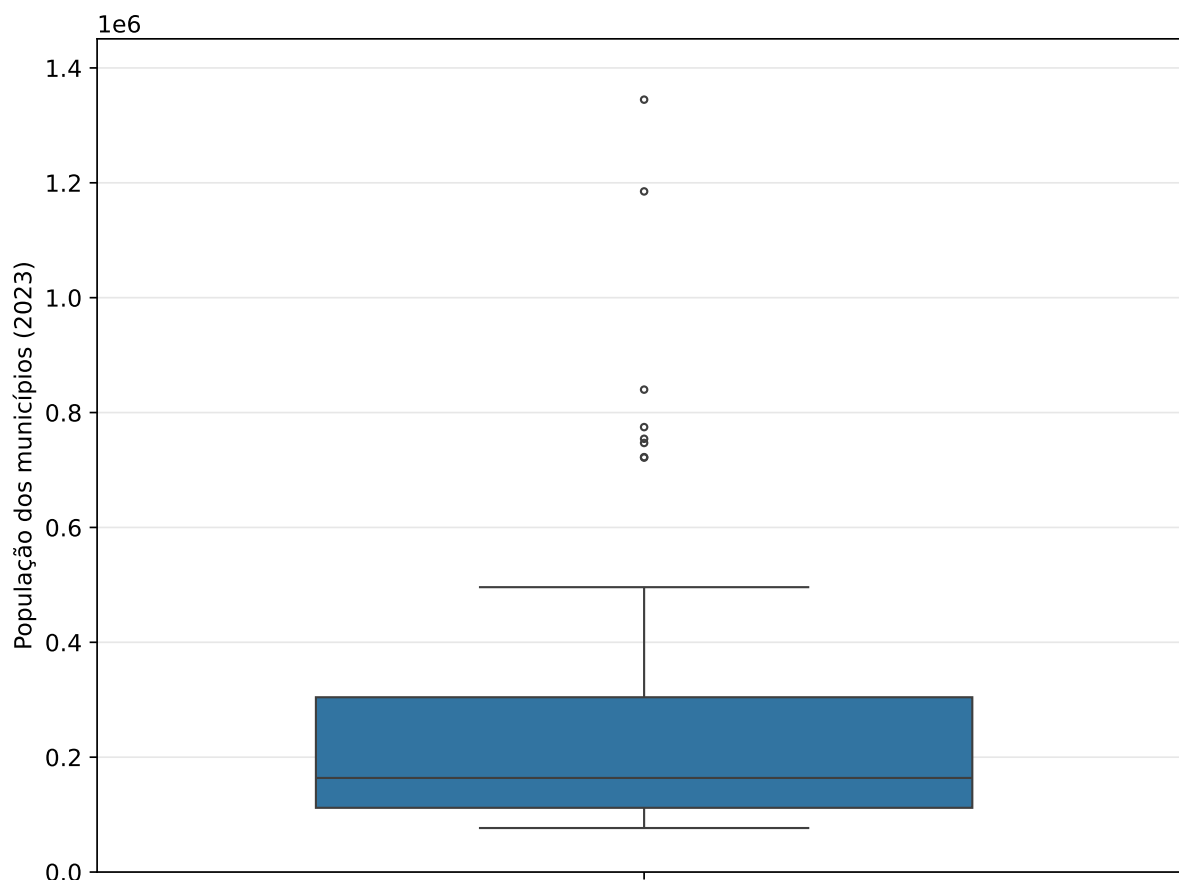
Durante os 24 anos, foram 116.160 óbitos de homens e 7.508 de mulheres, com 57 mortes sem identificação de sexo. Portanto, uma razão de mais de 15 óbitos de homens para cada óbito de mulher. O maior número de óbitos ocorreu nas faixas etárias de 20 a 24 anos (27.048), 25 a 29 anos (21.847) e 15 a 19 anos (20.059). Isso significa que 56,7% dos óbitos por armas de fogo ocorreram entre jovens, isto é, pessoas com 15 a 29 anos de idade.

Os óbitos por armas de fogo foram distribuídos segundo suas causas. No período, ocorreram 106.572 (86,1%) homicídios, 7.920 (6,4%) óbitos com causa indeterminada, 4.827 (3,9%) suicídios, 3.668 (3,0%) óbitos por intervenção legal e 738 (0,6%) óbitos acidentais.

Quanto aos cem municípios mais populosos do Estado de São Paulo, as populações no último ano da série variaram entre a capital, São Paulo, com 11.930.850 habitantes, e Lins, com 76.646 habitantes. A média das populações municipais foi de 366.938 habitantes, o desvio-padrão, de 1.189.748 habitantes e a mediana, de 163.826 habitantes.

O boxplot na Figura 1 demonstra a distribuição das populações municipais no ano de 2023, com truncamento no percentil 99, isto é, excluindo o município de São Paulo para evitar distorção no gráfico (São Paulo é quase nove vezes maior que o segundo maior município, Guarulhos).

Figura 1 – Distribuição das populações dos cem municípios mais populosos de São Paulo em 2023 (até percentil 99).



Fonte: elaborada pelo autor.

O gráfico evidencia ampla variabilidade no tamanho populacional dos municípios analisados, com assimetria. A maioria das cidades apresenta populações substancialmente inferiores aos maiores municípios do Estado, como Guarulhos (1.344.768 de habitantes), Campinas (terceiro mais populoso, com 1.184.995 habitantes), São Bernardo do Campo (quarto, com 839.921 habitantes) e Santo André (quinto, com 774.653 habitantes).

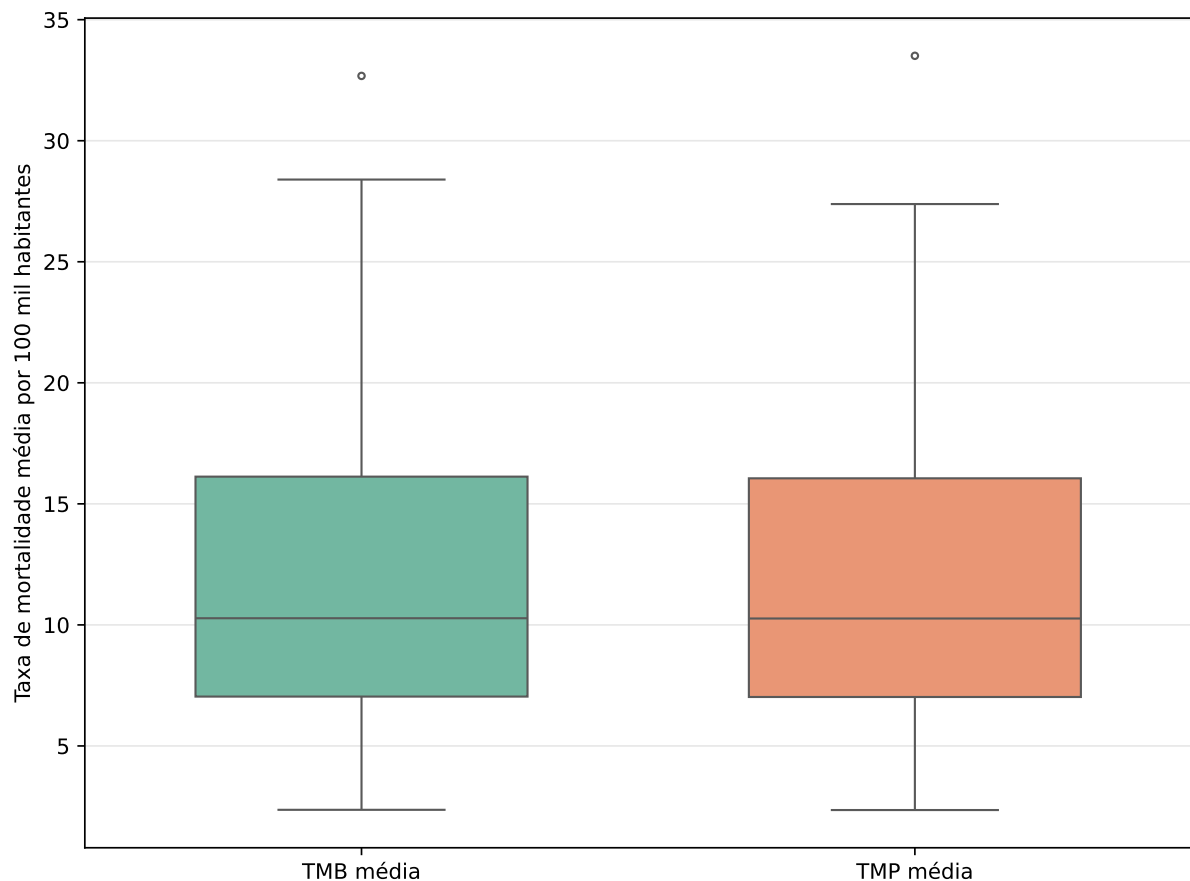
Os dados de mortalidade foram associados aos dados demográficos dos municípios paulistas, visando a criação de taxas de mortalidade por armas de fogo brutas (TMB) e padronizadas por idade e sexo (TMP) para os cem municípios mais populosos do Estado de São Paulo.

A Tabela 2 apresenta estatísticas descritivas dessas duas taxas anuais para os cem municípios analisados ao longo do período de 2000 a 2023. Portanto, são 2.400 instâncias de município-ano descritas pela quantidade de óbitos, pela população e pelas taxas bruta e padronizada de mortalidade por armas de fogo.

Tabela 2 – Estatísticas descritivas dos municípios-ano, referentes aos cem municípios mais populosos de São Paulo no período 2000-2023.

	Óbitos	População	Taxa bruta	Taxa padronizada
N	2400.00	2400.00	2400.00	2400.00
Média	45.86	340395.58	11.80	11.82
Desvio-padrão	210.04	1156801.56	11.24	11.09
Mínimo	0.00	48011.00	0.00	0.00
1º quartil	6.00	97783.50	4.61	4.70
Mediana	14.00	147367.00	8.47	8.54
3º quartil	31.00	277448.25	14.56	14.65
Máximo	4311.00	12015228.00	91.33	85.31

Figura 2 – Distribuição das taxas anuais médias de mortalidade bruta (TMB) e padronizada por idade e sexo (TMP) dos cem municípios mais populosos de São Paulo.



Fonte: elaborada pelo autor.

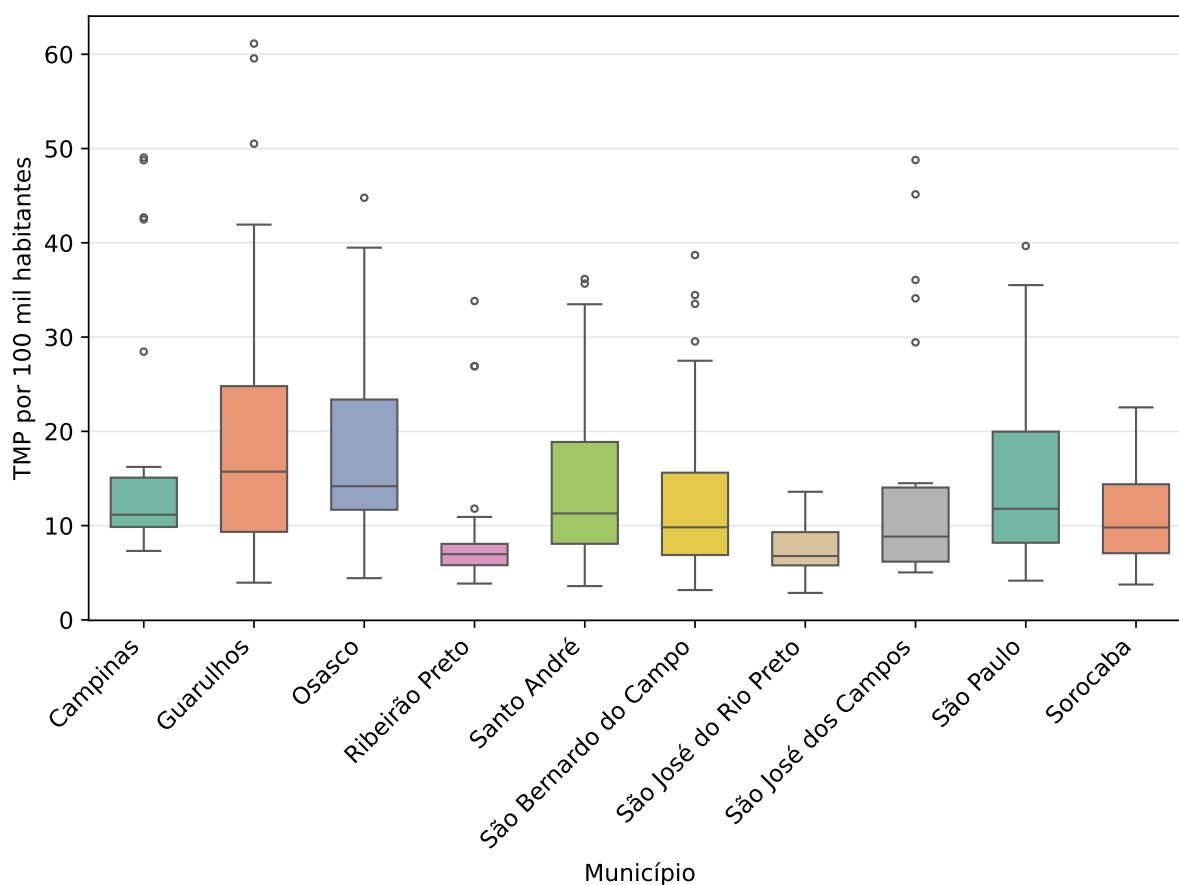
Nota-se que o número de óbitos por ano e município pode ser bastante baixo, incluindo nulos. A mediana é de 14 mortes, mas o último quartil dessa distribuição varia entre 31 e 4.311 óbitos. As taxas brutas e padronizadas apresentam diferenças discretas, conforme esperado. TMB e TMP possuem média de 11,8 óbitos por 100 mil habitantes e

mediana de 8,5 óbitos por 100 mil habitantes. A maior taxa bruta é de 91,3 óbitos por 100 mil habitantes, e a maior taxa padronizada é de 85,3 óbitos por 100 mil habitantes.

As séries anuais dos municípios foram sintetizadas em uma média referente aos 24 anos do período estudado. O boxplot na Figura 2 apresenta as duas taxas anuais médias para os cem municípios mais populosos, revelando que a padronização não muda muito em relação à taxa bruta, o que era esperado.

Outra forma de resumir o comportamento das taxas de mortalidade é observar seus valores por municípios. A Figura 3 apresenta boxplots das TMP segundo os dez municípios mais populosos do Estado.

Figura 3 – Distribuição das taxas de mortalidade padronizadas por idade e sexo (TMP) dos dez municípios mais populosos de São Paulo, 2000-2023.

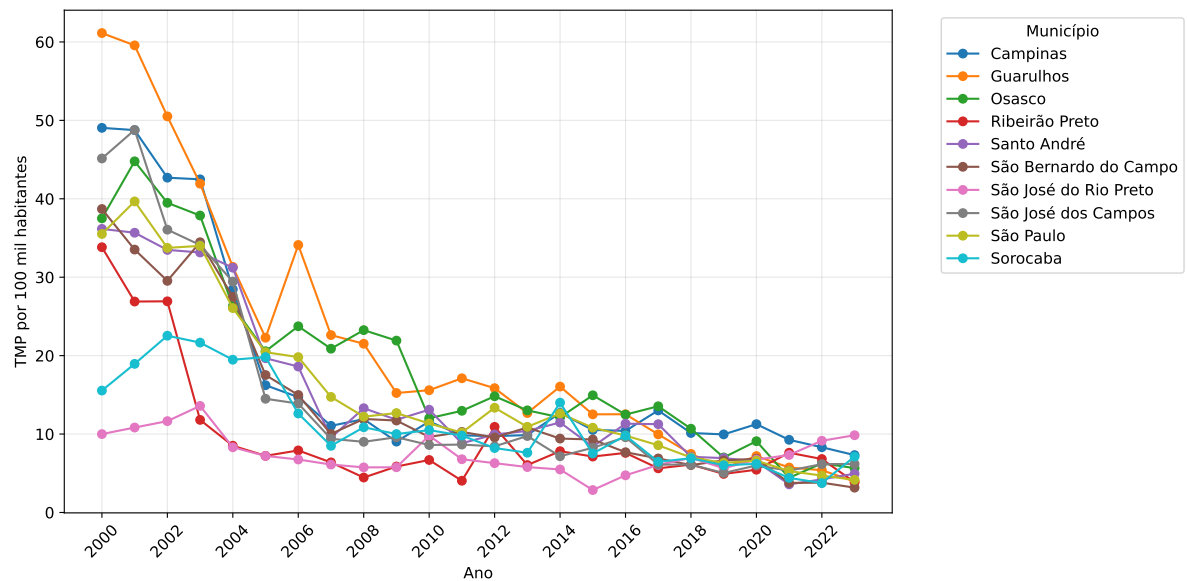


Fonte: elaborada pelo autor.

A princípio, observa-se que as taxas variam entre municípios em alguns aspectos, como concentração, patamar da mediana e quantidade de outliers. Por exemplo, São José do Rio Preto tem valores concentrados, enquanto Guarulhos possui maior dispersão; as medianas variaram entre Ribeirão Preto (7,0) e Guarulhos (15,7); São José dos Campos tem 5 valores aberrantes (outliers), enquanto Sorocaba não tem nenhum. Em geral, as distribuições das taxas municipais têm assimetria à direita.

Já na Figura 4, consta o gráfico das séries temporais das taxas padronizadas para os mesmos dez maiores municípios. Observa-se uma tendência de queda em geral para todas as taxas no período.

Figura 4 – Taxas de mortalidade padronizadas por idade e sexo (TMP) dos dez municípios mais populosos de São Paulo, 2000-2023.



Fonte: elaborada pelo autor.

## 4.2 Agrupamento de séries temporais

Em primeiro lugar, cabe reforçar que as análises de agrupamento realizadas a partir desta seção referem-se às taxas de mortalidade padronizadas por idade e sexo (TMP) dos cem municípios mais populosos do Estado. A TMP é a medida mais adequada para a comparação da mortalidade entre localidades e tempos distintos, pois corrige eventuais efeitos decorrentes das distintas estruturas populacionais dos municípios nos anos. Como descrito na metodologia, os agrupamentos aqui realizados utilizam Dynamic Time Warping (DTW) como medida de similaridade.

Inicialmente, procedemos com a elaboração de um heatmap com a matriz de distâncias utilizando DTW (Figura 5). Para tanto, utilizamos uma amostra menor (os dez municípios mais populosos), de forma a permitir uma boa visualização gráfica das distâncias.

Figura 5 – Heatmap da matriz de distâncias DTW dos dez municípios mais populosos de São Paulo.



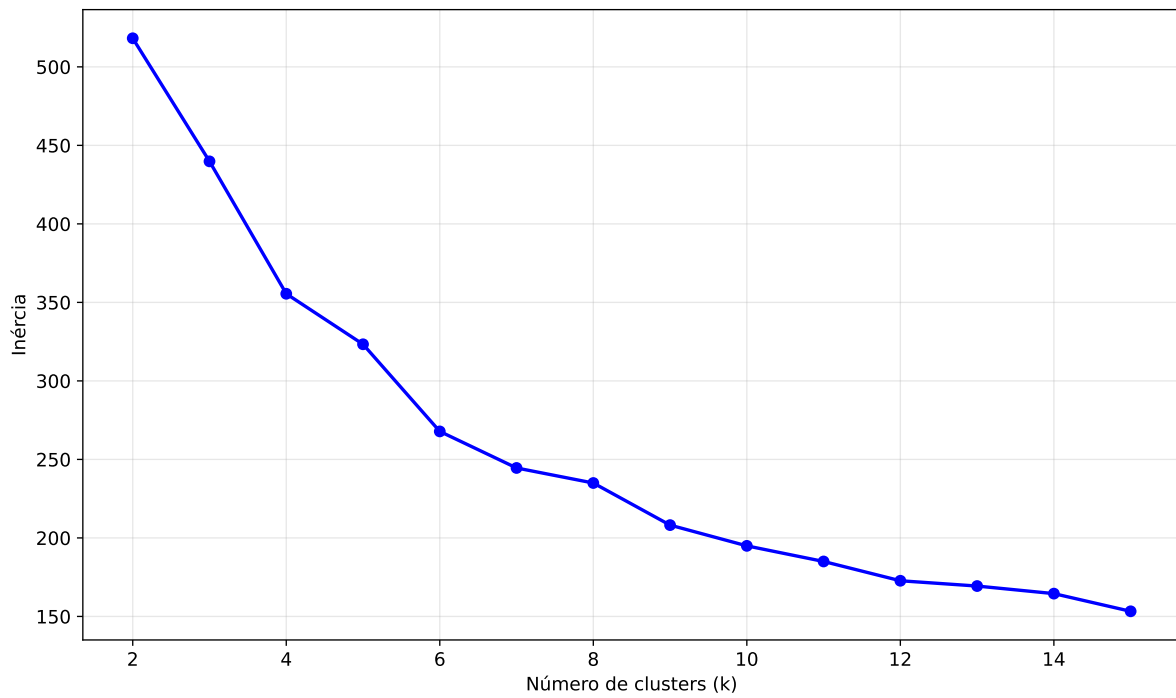
Fonte: elaborada pelo autor.

Nesta amostra menor, percebe-se como alguns municípios têm distâncias bastante grandes entre si, por exemplo nos pares: São José do Rio Preto e Guarulhos; Sorocaba e Guarulhos; São José do Rio Preto e Campinas. Outros municípios são mais próximos entre si, como São Paulo e outras cidades da mesma região metropolitana, tais como, São Bernardo do Campo, Santo André e Osasco. Essa métrica permite capturar a diferença visual observada entre as séries temporais na Figura 4.

Além disso, elaboramos um agrupamento hierárquico com o método de ligação por média dos grupos, resultando no dendrograma da Figura 6. O dendrograma permite observar com ainda mais facilidade como há diferenças e similaridades entre as séries

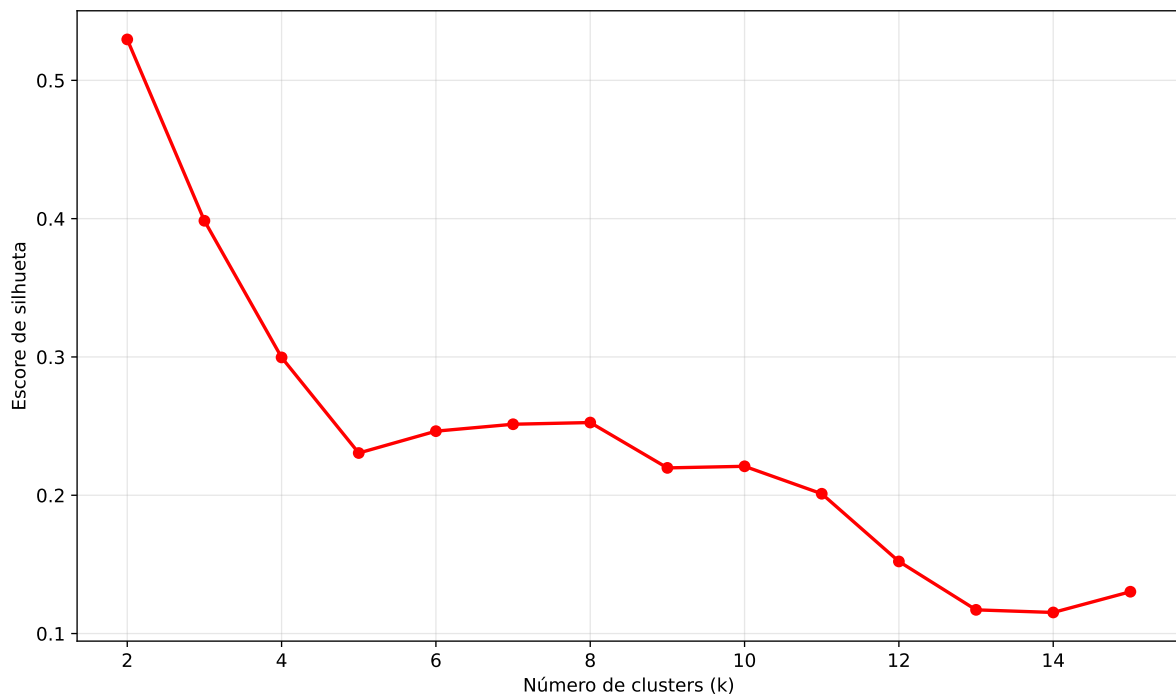


Figura 7 – Método do cotovelo (inércia).



Fonte: elaborada pelo autor.

Figura 8 – Escore de silhueta.

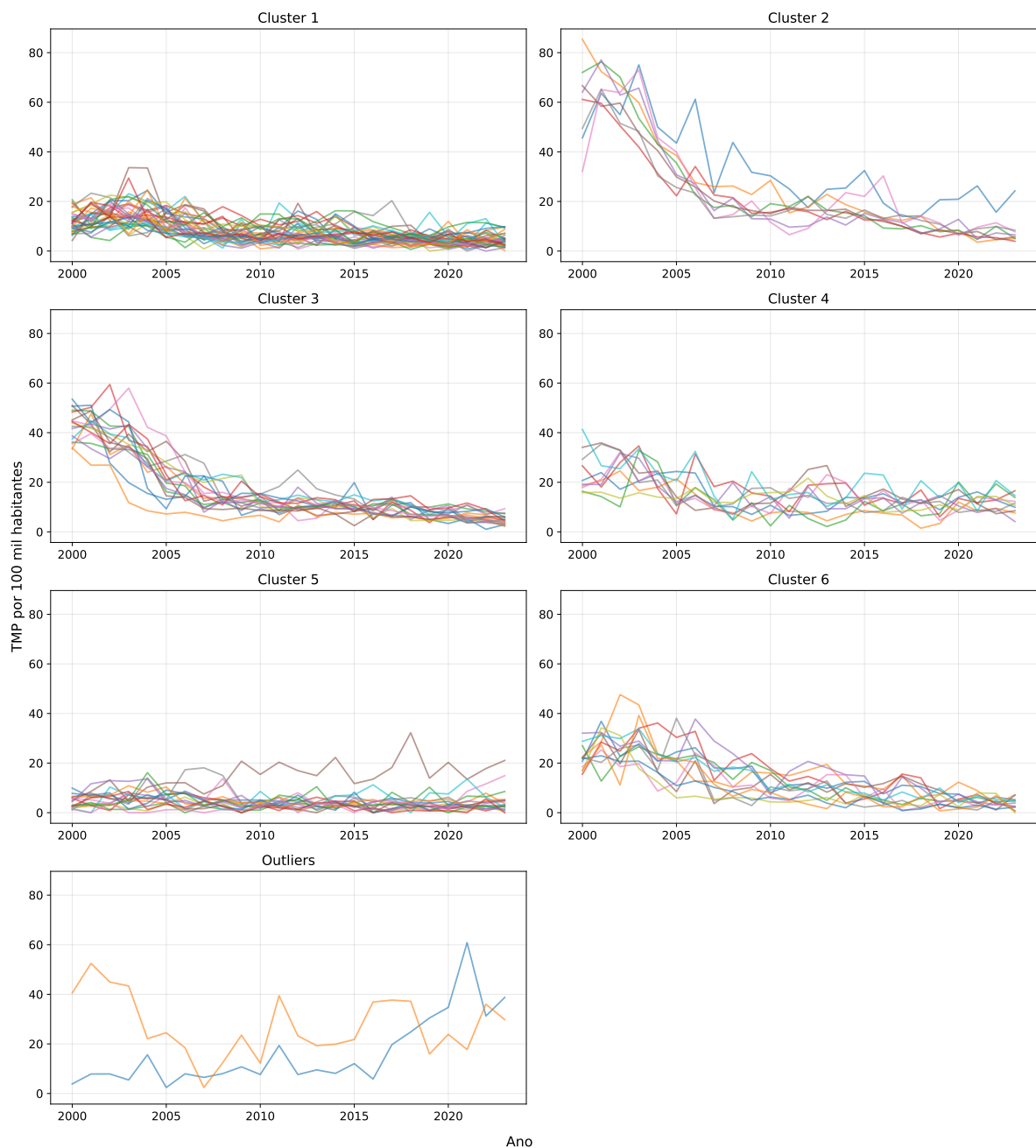


Fonte: elaborada pelo autor.

Utilizando o escore de silhueta e buscando o máximo local (Figura 8), o melhor K

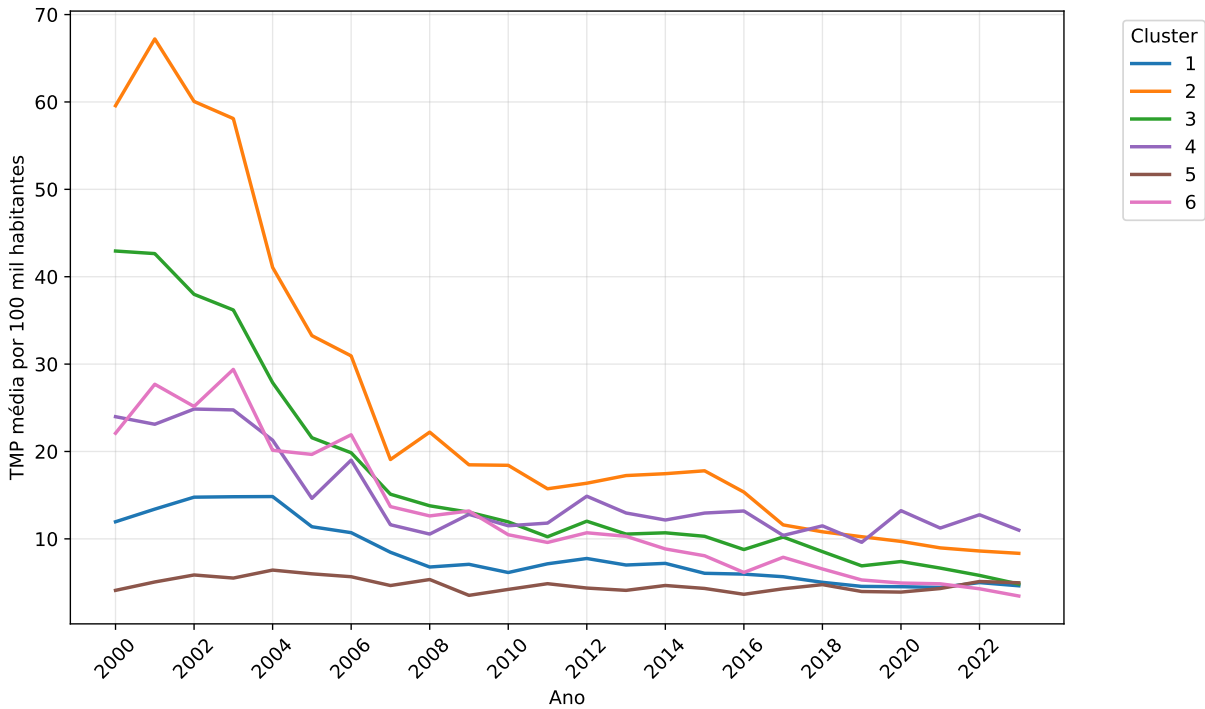
identificado foi para oito grupos (com um escore de 0,253). Ao agrupar as TMP nos oito clusters, foram identificados dois clusters com apenas um município cada. Esses municípios foram tratados como outliers conjunturais, pois suas curvas não se assemelham a nenhum outro caso da amostra. Assim, o modelo final teve seis agrupamentos e dois municípios não agrupados. As curvas individuais das taxas de mortalidade padronizadas de cada município nos seis clusters estão representadas na Figura 9.

Figura 9 – Taxas de mortalidade padronizadas por idade e sexo (TMP), 2000-2023. Curvas individuais de cada cluster com distância DTW.



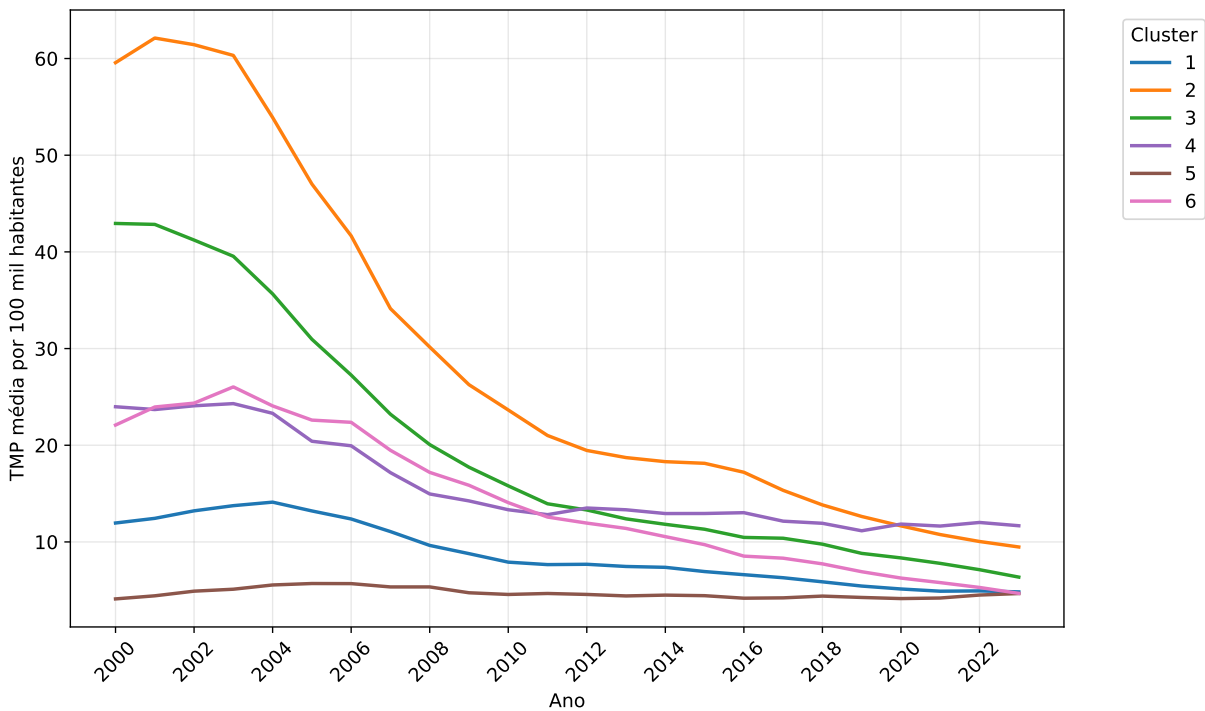
Fonte: elaborada pelo autor.

Figura 10 – Taxas médias de mortalidade padronizadas por idade e sexo (TMP média), 2000-2023. Curvas médias de cada cluster com distância DTW.



Fonte: elaborada pelo autor.

Figura 11 – Tendência suavizada (MMEP) das taxas médias de mortalidade padronizadas por idade e sexo (TMP média), 2000-2023. Curvas médias de cada cluster com distância DTW.



Fonte: elaborada pelo autor.

As taxas médias dos municípios que compõem cada grupo ajudam a resumir o comportamento das taxas individuais, conforme demonstrado na Figura 10.

Como não há padrão sazonal nas TMP anuais analisadas e com a intenção de visualizar a tendência das médias de cada cluster, utilizamos suavização por Média Móvel Exponencialmente Ponderada (MMEP) com uma janela de cinco anos, para destacar a tendência de longo prazo das taxas médias de mortalidade por armas de fogo de cada cluster (Figura 11).

A relação dos municípios referentes a cada cluster consta na Tabela 3.

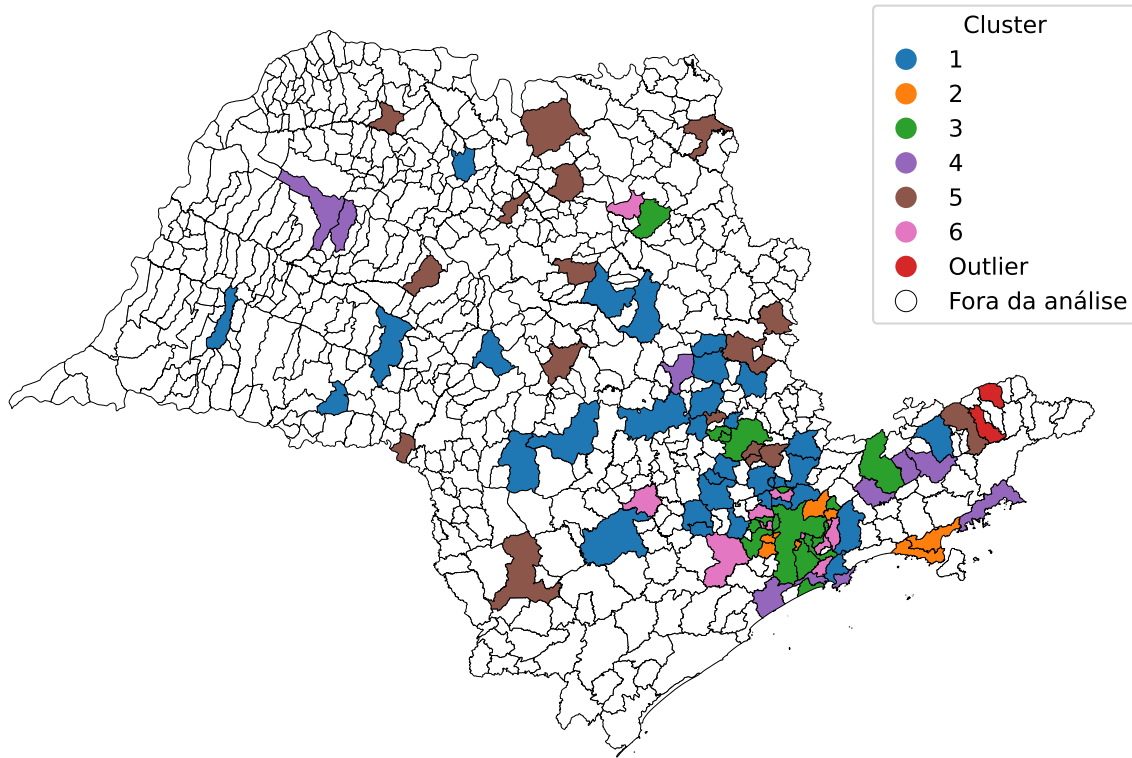
Tabela 3 – Municípios agrupados segundo clusters obtidos por K-Means com DTW.

Cluster	Total de municípios	Municípios
1	34	Araraquara, Araras, Assis, Atibaia, Avaré, Bauru, Botucatu, Bragança Paulista, Caieiras, Cajamar, Campo Limpo Paulista, Indaiatuba, Itapetininga, Itu, Jundiaí, Leme, Limeira, Mairiporã, Marília, Mogi Mirim, Mogi das Cruzes, Paulínia, Pindamonhangaba, Piracicaba, Presidente Prudente, Salto, Santa Bárbara d'Oeste, Santos, Sorocaba, São Carlos, São José do Rio Preto, São Roque, Votorantim, Várzea Paulista
2	8	Caraguatatuba, Diadema, Embu das Artes, Guarulhos, Itapeverica da Serra, Itaquaquecetuba, São Sebastião, Taboão da Serra
3	17	Arujá, Barueri, Campinas, Cotia, Ferraz de Vasconcelos, Francisco Morato, Hortolândia, Itapevi, Mauá, Osasco, Praia Grande, Ribeirão Preto, Santo André, Sumaré, São Bernardo do Campo, São José dos Campos, São Paulo
4	10	Araçatuba, Birigui, Caçapava, Guarujá, Itanhaém, Jacareí, Rio Claro, São Vicente, Taubaté, Ubatuba
5	17	Americana, Barretos, Bebedouro, Catanduva, Franca, Guaratinguetá, Itapeva, Itatiba, Jaú, Lins, Matão, Mogi Guaçu, Ourinhos, São João da Boa Vista, Valinhos, Vinhedo, Votuporanga
6	12	Carapicuíba, Cubatão, Franco da Rocha, Ibiúna, Jandira, Poá, Ribeirão Pires, Santana de Parnaíba, Sertãozinho, Suzano, São Caetano do Sul, Tatuí
Outlier	2	Cruzeiro, Lorena

Um mapa do Estado de São Paulo (Figura 12) foi elaborado com os agrupamentos resultantes dessa análise, de modo a visualizar se há alguma identidade espacial entre os

clusters das séries temporais das taxas de mortalidade por armas de fogo.

Figura 12 – Clusters de mortalidade por armas de fogo. Cem municípios mais populosos do Estado de São Paulo.



Fonte: elaborada pelo autor.

As séries temporais médias de cada cluster foram submetidas a um teste de estacionariedade (Teste de Dickey-Fuller Aumentado - ADF). Com um nível de significância de 0,05, as séries dos clusters 1 e 4 foram consideradas estacionárias, e as séries dos demais clusters foram consideradas não estacionárias.

### 4.3 Discussão sobre o agrupamento de séries temporais

O uso do algoritmo K-Means com DTW permitiu agrupar séries temporais municipais das taxas de mortalidade padronizadas por idade e sexo (TMP) de modo a maximizar a similaridade entre as séries do mesmo grupo e minimizar a similaridade entre os grupos. Portanto, algumas características comuns dentro de cada agrupamento podem ser observadas, considerando que na análise final obtivemos seis agrupamentos e dois municípios classificados como outliers.

O cluster 1 (cor azul nas Figuras 10, 11 e 12) agrega o maior número de municípios dentre todos os agrupamentos, 34. É caracterizado predominantemente por cidades do interior, contendo apenas quatro cidades da Região Metropolitana de São Paulo (RMSP). Há maior concentração de municípios da região de Campinas neste grupo. As TMP iniciam-se em um nível baixo em relação aos outros agrupamentos e aparentam ter uma queda bastante discreta no tempo. Esse padrão coincide com o esperado em cidades que já tinham menores índices de violência, o que costuma ocorrer em municípios do interior do Estado. Como a tendência geral dessa mortalidade no período é de queda, municípios que partem de um baixo patamar não teriam reduções muito expressivas em suas curvas.

Já o cluster 2 (cor laranja nas Figuras 10, 11 e 12) possui a queda mais nítida nas taxas de mortalidade por armas de fogo dentre todos os agrupamentos, inclusive por partir de um nível muito alto no início do período. A tendência mais acentuada de queda ocorre na primeira metade da série. Esse é o menor conjunto, com apenas oito municípios, formado por seis cidades da RMSP, tais como Guarulhos e Diadema, e por duas cidades do Litoral Norte. A região metropolitana é caracterizada por ser tradicionalmente um espaço mais violento do que o restante do Estado, o que poderia explicar esse patamar mais alto no início da série.

Outro agrupamento que contempla diversos municípios da Grande São Paulo é o cluster 3 (cor verde nas Figuras 10, 11 e 12), que possui 17 municípios. Além de 11 cidades da RMSP (incluindo a capital, São Paulo), conta com municípios da Região Metropolitana de Campinas (incluindo Campinas), uma cidade da Baixada Santista (Praia Grande) e outros dois grandes municípios do interior: Ribeirão Preto e São José dos Campos. Esse conjunto é caracterizado, portanto, por cidades com alto nível de urbanismo, o que costuma estar correlacionado a maior violência e maior frequência de crimes contra a vida em relação aos pequenos e médios municípios do interior. Nesse sentido, possui curvas com formato semelhante ao do cluster 2: taxas mais altas no início da série e queda visualmente identificável no tempo, em particular nos primeiros anos.

O cluster 3 diferencia-se do cluster 2 por possuir taxas iniciais em menor nível. No entanto, em ambos agrupamentos é nítida a redução das taxas no tempo e podemos compreender que ambos os clusters tem características municipais semelhantes, como

maior urbanismo, maior população e, por suposição, maior crime e maior disponibilidade de armas de fogo em relação aos municípios dos outros clusters.

Outro cluster com comportamento semelhante aos dois anteriores é o 6 (cor rosa nas Figuras 10, 11 e 12). Com 12 municípios, sendo oito pertencentes à RMSP e quatro cidades médias das Regiões de Sorocaba, Ribeirão Preto e Baixada Santista, observa-se também nítida queda nas curvas das TMP. Parte de um nível inicial de série em patamar mais baixo do que os clusters 2 e 3 – mas também chega a um nível, ao final da série, mais baixo do que esses dois grupos (é a menor TMP média entre todos os clusters no fim do período). Como se tratam de municípios da RMSP e de cidades mais desenvolvidas, localizadas perto de grandes centros regionais, as hipóteses de maior violência urbana e de mais disponibilidade de armas de fogo também podem ser fatores que atingem esses municípios.

Esses três conjuntos de cidades (clusters 2, 3 e 6) também podem representar municípios mais beneficiados pela política de recolhimento de armas de fogo de meados dos anos 2000, decorrente do Estatuto do Desarmamento. Como discutido na revisão bibliográfica, essa legislação é vista, por alguns autores, como um fator importante para a redução nas mortes por esse meio. Esses três clusters contam com 25 dos 29 municípios da RMSP que entraram na amostra (86,2%), além de centros regionais populosos e urbanizados do interior, como Campinas, Ribeirão Preto e São José dos Campos. Seria razoável imaginar que uma maior coleta de armas tenha ocorrido em cidades maiores, ainda que não tenhamos dados detalhados sobre a retirada de circulação de armas de fogo por cidades. Considerando os dados deste trabalho, observamos que as quedas mais nítidas nas taxas de mortalidade ocorreram nos anos 2000, período em que se deram as campanhas de desarmamento decorrentes dessa legislação.

O cluster 4 (cor roxo nas Figuras 10, 11 e 12) agrega dez municípios paulistas, quatro pertencentes à região do Vale do Paraíba e Litoral Norte, três da Baixada Santista e três do interior (regiões de Araçatuba e Campinas). Contém cidades com populações expressivas, como São Vicente, Taubaté e Guarujá. Trata-se de um cluster com um nível inicial semelhante ao do cluster 6, mas com uma redução nas taxas muito mais contida. De fato, ao fim do período estudado, é o agrupamento com maior taxa média dentre todos. Portanto, são cidades que não foram tão eficazes em reduzir a mortalidade por armas de fogo no tempo como os outros municípios paulistas conseguiram.

Por fim, o cluster 5 (cor marrom nas Figuras 10, 11 e 12) possui 17 municípios, nenhum pertencente à RMSP (assim como o cluster 4). Agrega municípios do interior de diversas regiões, semelhante ao que ocorre no cluster 1, mas com a diferença de não formar uma concentração em alguma região do Estado. São municípios, em geral, menos populosos. Esse cluster parte do nível mais baixo de mortalidade por armas de fogo dentre todos os agrupamentos e segue com taxas muito baixas em todo o período, sendo o conjunto

mais estável (sem tendência visual aparente) e com menor magnitude. Considerando ser composto por municípios menores do interior, esse comportamento parece condizente com localidades com menos violência urbana, em que o nível de mortalidade por armas de fogo seria igualmente baixo desde o início da série e, com isso, teria pouca margem para reduções significativas em suas curvas.

Duas séries temporais municipais não se assemelharam a nenhuma outra, formaram clusters unitários e, portanto, foram classificadas como outliers conjunturais. Foram as TMP dos municípios de Cruzeiro (cor azul no gráfico dos outliers da Figura 9) e de Lorena (cor laranja no gráfico dos outliers da Figura 9), ambas cidades do Vale do Paraíba. No primeiro caso, observamos um aumento na mortalidade por armas de fogo no fim do período e, em Lorena, há certa estabilidade, mas com muita flutuação. Essas duas séries foram interpretadas como trajetórias atípicas neste trabalho.

Particularmente, no caso de Cruzeiro, a curva é tão diferente das demais que já havia sido identificada assim no agrupamento hierárquico (Figura 6). A taxa aparenta estabilidade (ou leve crescimento) nos primeiros anos, mas apresenta acentuado crescimento a partir da segunda metade dos anos 2010. Assim, comporta-se como um caso extremo, com eventual crescimento da mortalidade por armas de fogo em um cenário de queda ou estabilização dessas mortes no conjunto dos outros municípios paulistas.

Chama a atenção que ambos os outliers pertencem à área do Vale do Paraíba (Região Administrativa de São José dos Campos). Importante relembrar que o cluster 4, cujas taxas ao fim do período são mais altas, em média, do que os outros agrupamentos, é aquele com maior número de cidades dessa região. Nesse sentido, esse território poderia ser objeto de uma investigação específica acerca do comportamento da mortalidade por armas de fogo.

Em resumo, há uma tendência geral de declínio nas taxas de mortalidade por armas de fogo no período, mas com diferenças relevantes entre os clusters. Nos agrupamentos 2, 3 e 6, essas quedas são claras, com mais força e com mais persistência do que nos clusters 1 e 4. Já no cluster 5, não há clareza sobre tendência, parecendo haver estagnação desses óbitos em um mesmo nível.

Complementando a análise de tendência gráfica, observamos a estacionariedade das séries médias dos clusters. Os clusters 2, 3, 5 e 6 tiveram suas curvas médias identificadas como não estacionárias por meio do Teste de Dickey-Fuller Aumentado. Os clusters 1 e 4 tiveram suas curvas médias identificadas como estacionárias, ou seja, média, variância e autocorrelação seriam contantes no tempo. Desse modo, as curvas médias dos clusters 1 e 4 teriam variações que representam apenas flutuações temporárias, isto é, os efeitos dos choques nas séries seriam transitórios e se dissipariam ao longo do tempo, fazendo com que a série retorne ao seu padrão médio. Já para as curvas médias dos outros quatro clusters, os choques que as séries sofrem teriam um efeito permanente sobre o seu nível – a

média ou a variância deixariam de ser constantes e as variações ocorridas em determinado período se acumulam e mudam a estrutura da curva de mortalidade.

Um último ponto fundamental para essa discussão diz respeito aos dados sobre disponibilidade de armas de fogo e outras informações sobre comércio e porte dessas armas no Brasil. Não temos boas fontes de dados que permitam estimar a quantidade de armas de fogo disponíveis pela população, o que dificulta criar associações entre a quantidade ou a densidade desses dispositivos em um município ou região e a quantidade de óbitos decorrentes do disparo de armas de fogo na mesma área. Portanto, é pertinente realizar estudos como este, com delineamento ecológico de séries temporais, pois apoiam a identificação do fenômeno da violência armada letal por diferentes prismas, ainda que não se possa falar em associação com outros fatores (como a disponibilidade de armas) – e muito menos em causalidade. Neste trabalho, vimos que o fenômeno da redução das taxas de mortalidade por armas de fogo não foi homogêneo entre os municípios paulistas.

## 5 CONCLUSÃO

Este trabalho buscou apresentar uma nova forma de abordar as séries temporais das taxas de mortalidade por armas de fogo, ao agrupar séries municipais segundo o algoritmo de aprendizado não supervisionado K-Means, utilizando a distância Dynamic Time Warping (DTW), medida de similaridade própria para séries temporais.

Para atingir esse objetivo principal, foram realizadas diversas operações em bancos de dados públicos, para chegar a bases de dados consistentes que permitissem cumprir com o objetivo. A localização das bases públicas corretas no ambiente virtual do Datasus, a filtragem necessária para local e tempo, bem como todo o processamento de variáveis, levaram à criação de alguns bancos de dados limpos e suficientemente completos para proceder com a análise exploratória, a criação de taxas metodologicamente adequadas e as análises com algoritmos de ciência de dados.

A análise exploratória de dados demonstrou o volume de óbitos por armas de fogo no Estado de São Paulo, a grande disparidade na ocorrência entre homens e mulheres (mais de 15 vezes), sua frequência atingindo os mais jovens e a predominância dos homicídios em relação às outras formas de violência. Também apontou que o número de óbitos por ano e município pode variar bastante, havendo municípios-ano sem mortes por essa causa e aqueles com mais de 4.000 óbitos. As taxas de mortalidade variam entre os municípios, mas em geral observa-se uma tendência de queda no período estudado de 24 anos.

Na análise de agrupamento, foram identificados municípios com taxas de mortalidade padronizadas por idade e sexo (TMP) semelhantes entre si. Observa-se que a redução na mortalidade por armas de fogo não foi homogênea entre os municípios do Estado de São Paulo. Assim, pode-se visualizar tendências e outros comportamentos a respeito dessa morte evitável em cada conjunto, capacitando gestores públicos a proceder com medidas focalizadas para alguns municípios e com políticas públicas estaduais de prevenção à violência letal melhor direcionadas.

Há limitações nas análises elaboradas no âmbito deste projeto, tais como: a definição de hiperparâmetros do modelo (como o número de clusters) impacta diretamente na definição de qual município é considerado semelhante ou diferente de outro município; a classificação de óbito por arma de fogo pode ter viés de registro, e esse viés ser distribuído de modo desigual entre as diferentes regiões ou municípios do Estado; a performance do teste de estacionariedade (ADF) pode ter sido afetada pelo baixo número de observações das séries temporais (24 anos), reduzindo o poder do teste.

Contudo, este tipo de análise permite ser um ponto de partida para estudos futuros. Por exemplo, desenvolver modelos para identificar e quantificar a tendência da mortalidade

por armas de fogo de cada agrupamento ou modelos preditivos das taxas futuras de cada cluster. Análises mais individualizadas também são possíveis, como aprofundar a investigação sobre os outliers identificados ou sobre a região a que eles pertencem. O trabalho também pode inspirar a elaboração de estudos de agrupamento para outros fenômenos sociais, econômicos ou epidemiológicos que são descritos por meio de séries temporais, tais como a análise de outras taxas de mortalidade, a ocorrência de crimes, a incidência de doenças de interesse da saúde pública ou mesmo a avaliação de valores referentes à arrecadação tributária. O uso de algoritmos de agrupamento, como o realizado no âmbito deste estudo para identificar séries temporais de mortalidade por armas de fogo semelhantes entre si e que se diferenciam de outras, permite apoiar o desenvolvimento e a avaliação de políticas públicas diversas.

## REFERÊNCIAS

- ALVES, L. G.; RIBEIRO, H. V.; RODRIGUES, F. A. Crime prediction through urban metrics and statistical learning. **Physica A: Statistical Mechanics and its Applications**, Elsevier, v. 505, p. 435–443, 2018.
- BEYAZTAS, U.; SHANG, H. Machine-learning-based functional time series forecasting: application to age-specific mortality rates. **Forecasting**, MDPI, v. 4, n. 1, p. 394–408, 2022.
- BRASIL. MINISTÉRIO DA SAÚDE. **Mortes por causas externas: qualificação dos registros inespecíficos**. Brasília: Ministério da Saúde, 2024. 113 p. ISBN 9786559935765.
- DUARTE, T. S. B. **Mortalidade por suicídio na cidade de São Paulo: caracterização, tendências e análise de efeito idade-período-coorte (1996-2018)**. 2022. Dissertação (Mestrado) — Faculdade de Saúde Pública, Universidade de São Paulo, 2022.
- GOLDSTEIN, E. V.; WILSON, F. A. Predicting state-level firearm suicide rates: a machine learning approach using public policy data. **American journal of preventive medicine**, Elsevier, v. 67, n. 5, p. 753–758, 2024.
- JORGE, M. A. Uma análise da mensuração da disponibilidade de armas de fogo no brasil. **Economic Analysis of Law Review**, v. 11, n. 2, p. 97–125, 2020.
- MALTA, D. C. *et al.* Association between firearms and mortality in brazil, 1990 to 2017: a global burden of disease brazil study. **Population health metrics**, Springer, v. 18, n. Suppl 1, p. 19, 2020.
- MCDONALD, K. *et al.* Trends in method-specific suicide in brazil from 2000 to 2017. **Social psychiatry and psychiatric epidemiology**, Springer, v. 56, p. 1779–1790, 2021.
- MEIRA, K. C. *et al.* Temporal trends in female firearm homicides across states in the northeast of brazil during the period 2000-2019. **Ciência & Saúde Coletiva**, SciELO Public Health, v. 29, p. e14892022, 2024.
- MURRAY, J.; CERQUEIRA, D. R. de C.; KAHN, T. Crime and violence in brazil: Systematic review of time trends, prevalence rates and risk factors. **Aggression and violent behavior**, Elsevier, v. 18, n. 5, p. 471–483, 2013.
- NIELSEN, A. **Análise prática de séries temporais: predição com estatística e aprendizado de máquina**. Rio de Janeiro: Alta Books, 2021.
- ORGANIZAÇÃO MUNDIAL DA SAÚDE. **Guia de Referência da CID-11**. Genebra: Organização Mundial da Saúde, 2022.
- PERES, M. F. T.; SANTOS, P. C. d. Mortalidade por homicídios no brasil na década de 90: o papel das armas de fogo. **Revista de saúde pública**, SciELO Public Health, v. 39, p. 58–66, 2005.
- PERLA, F. *et al.* Time-series forecasting of mortality rates using deep learning. **Scandinavian Actuarial Journal**, Taylor & Francis, v. 2021, n. 7, p. 572–598, 2021.

POLCARI, A. M. *et al.* A novel machine-learning tool to identify community risk for firearm violence: the firearm violence vulnerability index. **Journal of trauma and acute care surgery**, LWW, v. 95, n. 1, p. 128–136, 2023.

PORTA, M. *et al.* **A Dictionary of Epidemiology**. Sixth edition. New York: Oxford University Press, 2014. 343 p. (Oxford Reference). ISBN 9780199976720.

RIBEIRO, J. *et al.* Prediction of homicides in urban centers: A machine learning approach. *In*: SPRINGER. **Proceedings of SAI Intelligent Systems Conference**. Cham: Springer International Publishing, 2021. p. 344–361.

RODRIGUES, C. D. *et al.* Trends in suicide rates in brazil from 1997 to 2015. **Brazilian Journal of Psychiatry**, SciELO Brasil, v. 41, p. 380–388, 2019.

SANTIAGO, M. L. de O.; NUNES, R. A. de L.; MACENA, R. H. M. Tendência temporal dos homicídios no brasil no período de 2000-2019. **Journal of Health & Biological Sciences**, v. 9, n. 1, p. 1–10, 2021.

SOARES, A. de A. *et al.* Mortalidade de mulheres por arma de fogo no brasil: Análise temporal (2010-2019). **Cadernos Cajuína**, v. 9, n. 4, p. e249434–e249434, 2024.

SWEDO, E. A. *et al.* Development of a machine learning model to estimate us firearm homicides in near real time. **JAMA network open**, American Medical Association, v. 6, n. 3, p. e233413–e233413, 2023.

TAVENARD, R. *et al.* Tslern, a machine learning toolkit for time series data. **Journal of machine learning research**, v. 21, n. 118, p. 1–6, 2020.

UNITED NATIONS OFFICE ON DRUGS AND CRIME. **Global Study on Homicide 2023**. New York: UNODC, 2023. 162 p. ISBN 9789210030175.