

Estudo de caso: Suicídios na época da pandemia COVID-19

David Batista da Silva Gabriel Alves dos Santos
Maria Eduarda Bovo

Dezembro de 2025

Resumo

Este trabalho é uma análise da incidência de suicídio durante e após a pandemia COVID-19, compreendendo o período de 2020 a 2024. Para tanto, foi feito uso de técnicas de mineração de dados como o algoritmo k-means e de agrupamento, regras de associação e dimensionamento dos clusters usando a base de dados DATASUS.

Abstract

This work is an analysis of the incidence of suicide during and after the COVID-19 pandemic, covering the period from 2020 to 2024. To this end, data mining techniques such as the k-means algorithm and clustering, association rules, and cluster sizing were used, employing the DATASUS database.

1 Introdução

O suicídio no Brasil é um problema delicado que exige ações coordenadas de autoridades e profissionais de saúde. A urgência deste estudo é motivada por:

- Estima-se um aumento de **25% em ansiedade e depressão** pós-pandemia (OPAS/OMS, 2022), agravando a situação de saúde pública[cite: 6].
- Necessidade de **geração de conhecimento não explícito** sobre o tema a partir da base de dados DATASUS.

O projeto buscou investigar padrões de mortalidade por **lesões autoprovocadas intencionalmente** (CID-10: X60-X84) e identificar perfis de risco e padrões ocultos por meio de **clustering** para subsidiar políticas públicas.

2 Dados e Tratamento

2.1 Fonte de Dados

O Ministério da Saúde disponibiliza o **DATASUS** como o principal portal de dados do SUS para consulta pública[cite: 10]. Os dados de mortalidade são definidos pela Classificação Internacional de Doenças (CID-10), especificamente os códigos **X60 a X84** (Lesões Autoprovocadas Intencionalmente). O dataset original continha 17 atributos, incluindo raça, causa, estado civil e ano, com dados faltantes, categóricos e numéricos.

	ESTADO	ANO	DTOBITO	NATURAL	DTNASC	IDADE	SEXO	RACACOR	ESTCIV	ESC	OCUP	CODMUNRES	LOCOCOR	CODMUNOCOR	CAUSABAS	ESC2010
0	PR	1996	1996-10-24	841.0	1963-08-05	33.0	masculino	NaN	Solteiro	Nenhuma	61200.0	Pinhal de São Bento	outros	Pinhal de São Bento	X709	NaN
1	PR	1996	1996-02-04	152.0	1920-09-21	75.0	masculino	NaN	Casado	Nenhuma	62100.0	Assaf	domicilio	Assaf	X680	NaN
2	PR	1996	1996-01-17	841.0	1961-08-28	34.0	masculino	NaN	Solteiro	NaN	62100.0	Abatiá	hospital	Abatiá	X680	NaN
3	PR	1996	1996-02-07	841.0	1977-09-24	18.0	masculino	NaN	Solteiro	Nenhuma	62100.0	Santa Cecília do Pavão	hospital	São Sebastião da Amoreira	X680	NaN
4	PR	1996	1996-02-09	841.0	1956-08-30	39.0	masculino	NaN	União consensual	NaN	62100.0	Santa Mariana	domicilio	Santa Mariana	X700	NaN
5	PR	1996	1996-02-12	841.0	1970-12-04	25.0	masculino	NaN	Solteiro	NaN	62100.0	Santa Mariana	domicilio	Santa Mariana	X700	NaN
6	PR	1996	1996-01-04	841.0	1973-04-15	22.0	masculino	NaN	Solteiro	8 a 11 anos	39300.0	Cornélio Procópio	hospital	Cornélio Procópio	X628	NaN

Figure 1: Base de dados bruta.

2.2 Tratamento dos Dados

O tratamento dos dados focou na preparação do dataset para os algoritmos de mineração:

- **Remoção de colunas:** Foram removidas colunas que não apresentavam ganho de conhecimento para a análise ('ESC2010', 'ESCFALAGR1', 'CODMUNRES', 'LOCOCOR', 'NATURAL', 'CODMUNOCOR', 'OCUP').
- **Filtro Temporal:** O dataset foi filtrado para registros a partir de 2020.

- **Limpeza de Nulos:** Todos os registros com valores nulos foram removidos do dataset final de análise.

	Unnamed: 0	ESTADO	ANO	DTOBITO	IDADE	SEXO	RACACOR	ESTCIV	ESC	OCUP	CAUSABAS	MES OBITO
0	2889	SP	1996	1996-12-11	76.0	masculino	Parda	Viúvo	Nenhuma	58300.0	X740	12
1	2893	SP	1996	1996-12-26	32.0	masculino	Branca	Casado	12 anos e mais	31400.0	X794	12
2	2919	SP	1996	1996-12-25	31.0	masculino	Amarela	Solteiro	8 a 11 anos	41000.0	X790	12
3	3122	SP	1996	1996-04-18	45.0	feminino	Branca	Solteiro	Nenhuma	41000.0	X848	4
4	3127	SP	1996	1996-04-15	70.0	masculino	Amarela	Casado	12 anos e mais	700.0	X740	4
5	3130	SP	1996	1996-04-09	65.0	masculino	Amarela	Casado	Nenhuma	700.0	X798	4
6	3134	SP	1996	1996-04-15	18.0	feminino	Amarela	Solteiro	8 a 11 anos	14000.0	X709	4
7	3137	SP	1996	1996-04-05	30.0	feminino	Branca	Solteiro	12 anos e mais	14000.0	X700	4
8	3141	SP	1996	1996-04-18	50.0	masculino	Amarela	Separado judicialmente	12 anos e mais	2100.0	X749	4
9	3157	SP	1996	1996-04-24	39.0	masculino	Amarela	Casado	12 anos e mais	2100.0	X740	4

Figure 2: Base de dados limpa (Exemplo 2).

3 Resultados e Análise

3.1 Análise Exploratória e Insights

Os gráficos a seguir fornecem insights iniciais sobre a distribuição dos óbitos.

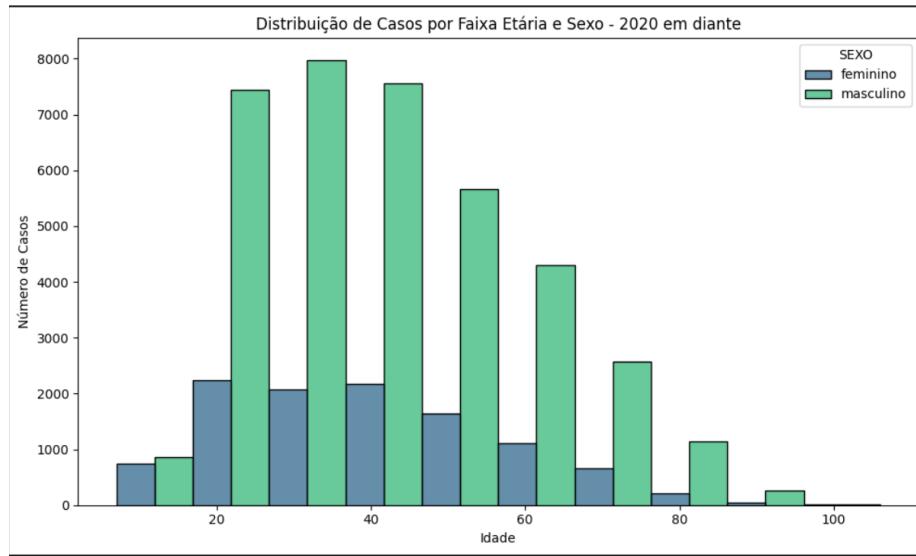


Figure 3: distribuição de Casos por Faixa Etária e Sexo - 2020 em diante. A maioria dos casos se concentra nas faixas etárias jovens e há uma predominância de casos no sexo masculino.

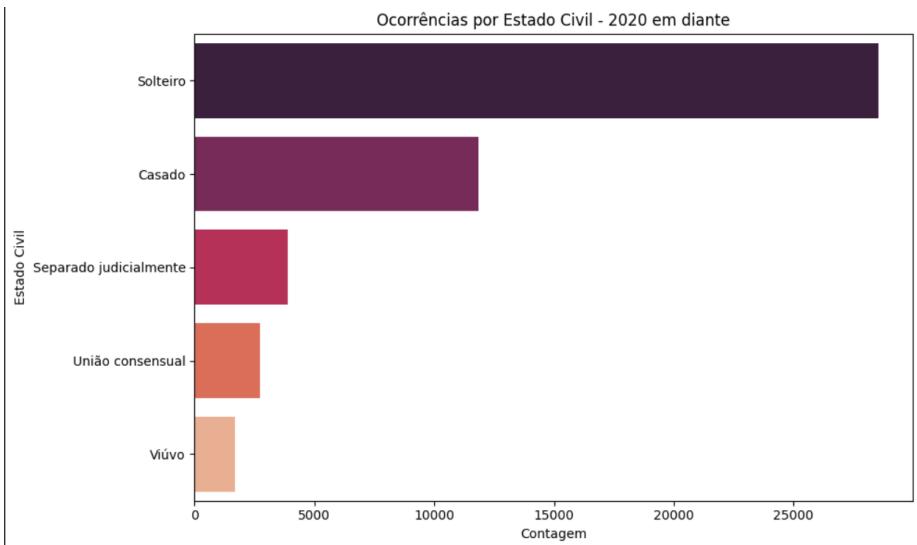


Figure 4: ocorrências por Estado Civil - 2020 em diante. O estado civil Solteiro é o que apresenta o maior número de ocorrências, seguido por Casado.

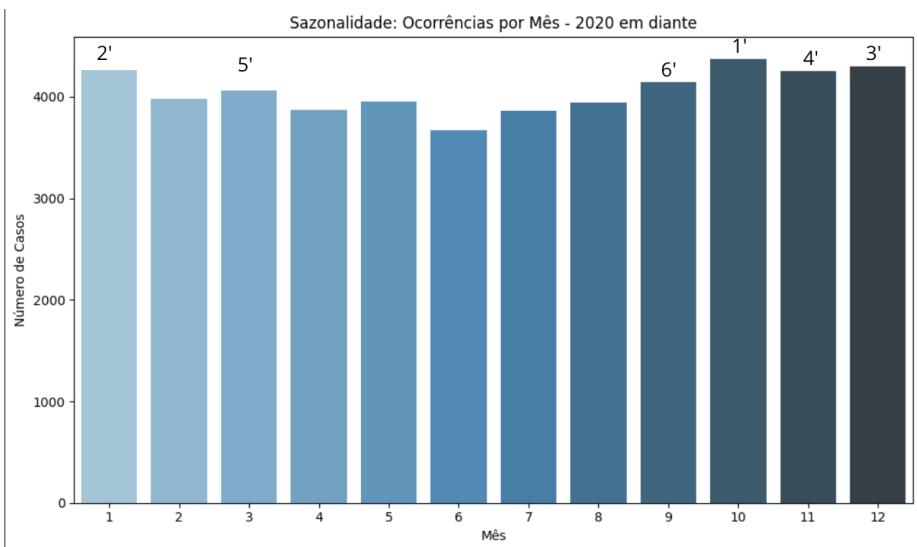


Figure 5: Sazonalidade: Ocorrências por Mês - 2020 em diante.

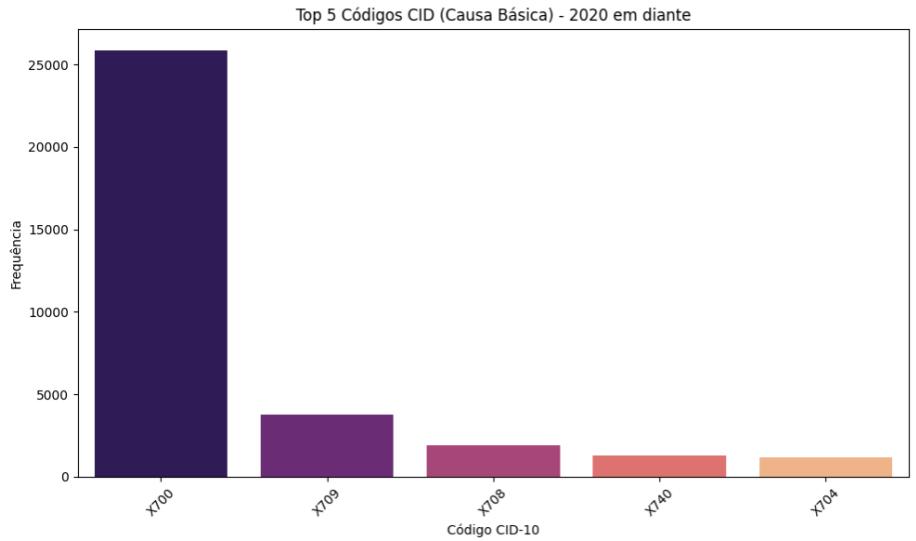


Figure 6: Top 5 Códigos CID (Causa Básica) - 2020 em diante: o X700 (Asfixia/Enforcamento em residência) é o mais frequente.

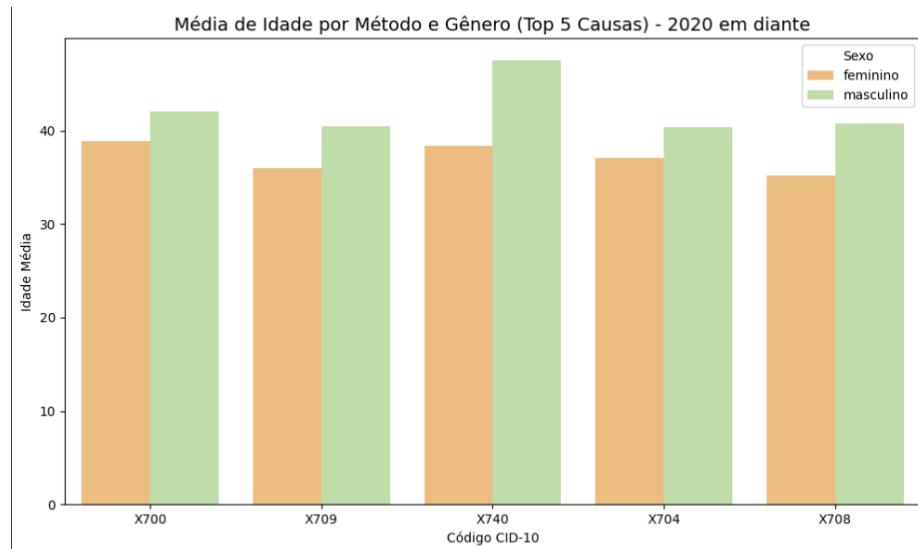


Figure 7: Média de Idade por Método e Gênero (Top 5 Causas) - 2020 em diante.

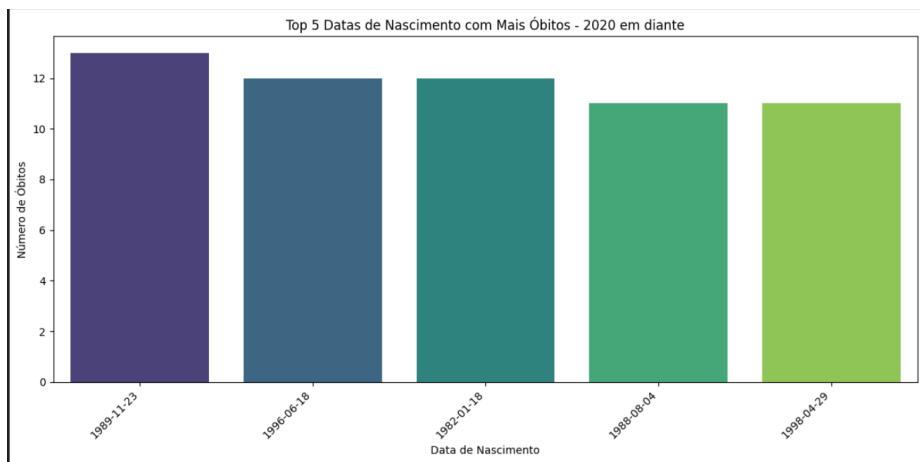


Figure 8: Top 5 Datas de Nascimento com Mais Óbitos - 2020 em diante.

3.2 Agrupamento (K-Means)

O Agrupamento (*Clustering*) é uma técnica de aprendizado não supervisionado que organiza objetos em grupos (clusters) baseados em similaridade.

Antes da aplicação do K-Means, os dados foram tratados por:

- **One-Hot Encoding:** Transformação de textos em números (ex: `Sexo_M` e `Sexo_F`).
- **Padronização (StandardScaler):** Colocação de todos os dados na mesma escala para equilibrar o peso das variáveis, impedindo que a "Idade" domine o cálculo da distância[cite: 88, 89].

O algoritmo *k-means* busca minimizar a função de custo J (Inércia), que é a soma dos quadrados das distâncias dos pontos aos centróides do seu cluster:

$$J = \min_C \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (1)$$

Escolha do K (Número de Clusters) Utilizamos o gráfico **Método Elbow (Cotovelo)** para decidir o número ideal de grupos, que é onde a linha do erro faz uma curva brusca e para de cair drasticamente.

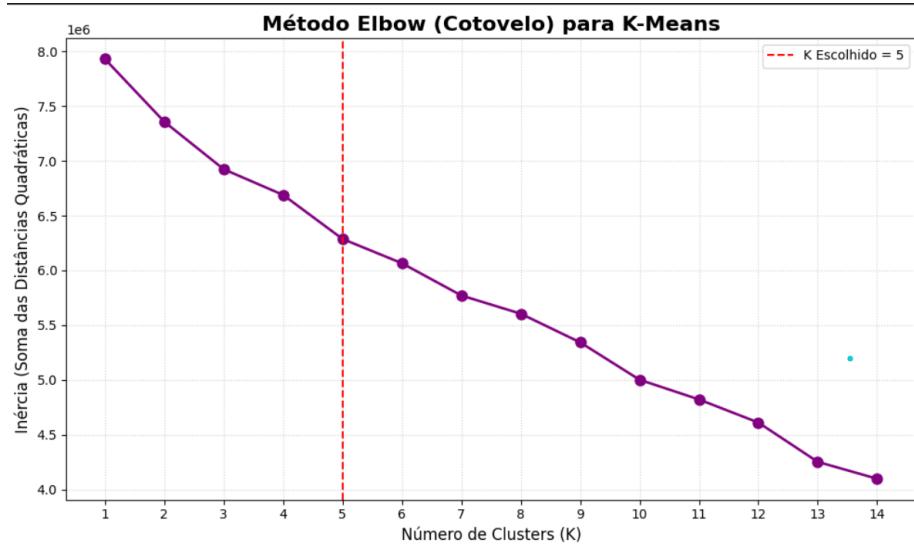


Figure 9: Método Elbow (Cotovelo) para K-Means. Escolha final: $\mathbf{K = 5}$ clusters, pois oferece a melhor divisão sem fragmentar demais os dados[cite: 93].

Perfis dos Clusters ($K=5$) O algoritmo identificou 5 perfis distintos de comportamento, sendo os Clusters 1 e 2 os mais representativos[cite: 94].

- **Cluster 1 (39.9% dos registros):** Idade Média: **33 anos**. Perfil: Predominantemente **Masculino**, **Solteiro**, Raça **Parda**, com escolaridade de 8 a 11 anos (Ensino Médio).
- **Cluster 2 (28.0% dos registros):** Idade Média: **53 anos**[cite: 98]. Perfil: Predominantemente **Masculino**, **Casado**, Raça **Branca**, com escolaridade baixa (4 a 7 anos). É o grupo com a maior idade média e o único onde o estado civil Casado é a moda.

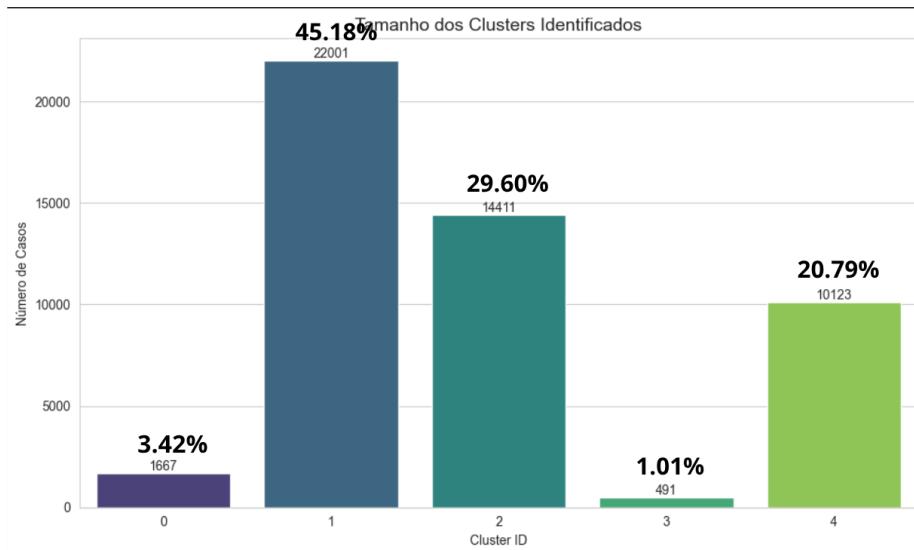


Figure 10: Tamanho dos Clusters Identificados.

3.3 Regras de Associação

As Regras de Associação (Algoritmo Apriori) foram usadas para identificar padrões de co-ocorrência entre variáveis demográficas e as causas de morte (CID-10). As variáveis foram categorizadas (ex: Idade em faixas de 15 anos) para criar transações de itens.

3.3.1 Dados Usados

Variável	Categorização	Exemplo
IDADE	Faixas de 15 anos	IDADE_45-59
SEXO	M/F	SEXO_M
RACACOR	Categorias IBGE	RACA_1
ESTCIV	Estado civil	ESTCIV_2
ESC	Nível educacional	ESC_3
CAUSABAS	Código CID-10	CAUSA_I21

Figure 11: dados para as regras de associação.

Filtros de Qualidade Para garantir a relevância estatística, foram utilizados filtros rigorosos:

- **Suporte** $\geq 3\%$
- **Confiança** $\geq 35\%$
- **Lift** ≥ 1.2

As métricas avaliadas são:

1. **Suporte (Sup):**

$$\text{Support}(A) = \frac{|\{t \in D \mid A \subseteq t\}|}{|D|} \quad (2)$$

2. **Confiança (Conf):**

$$\text{Confidence}(A \Rightarrow B) = \frac{\text{Support}(A \cup B)}{\text{Support}(A)} \quad (3)$$

3. **Lift:** (Mede a associação; $Lift > 1$ indica correlação positiva)

$$\text{Lift}(A \Rightarrow B) = \frac{\text{Confidence}(A \Rightarrow B)}{\text{Support}(B)} = \frac{\text{Support}(A \cup B)}{\text{Support}(A) \times \text{Support}(B)} \quad (4)$$

Principal Descoberta O padrão **Jovens (15-29 anos) e óbito por arma de fogo** tem uma associação **66% superior ao acaso** ($\text{Lift} \approx 1.66$) para este grupo etário.

- **35,6% dos óbitos** entre 15-29 anos são por arma de fogo.
- As regras mais fortes se concentraram na associação entre **Ensino Médio** → **Causa X700** e **Estado Civil Solteiro** (conforme a tabela de regras).

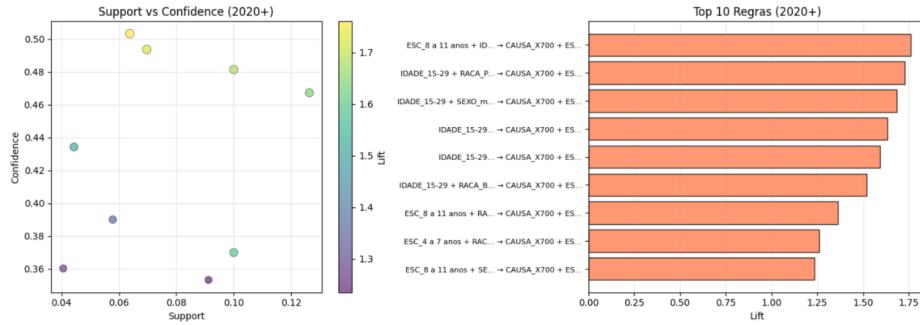


Figure 12: Top 10 Regras (2020+) e Gráfico Support vs Confidence. O Lift elevado indica forte associação entre características demográficas e a causa X700 (Asfixia/Enforcamento) + Estado Civil Solteiro.

4 Conclusão

O estudo, ao utilizar K-Means e Regras de Associação, forneceu evidências robustas para a criação de políticas de prevenção segmentadas.

- **Agravamento Pós-COVID:** O aumento de ansiedade/depressão reforça a urgência do tema.
- **Perfis de Risco Claros:** Identificação de dois perfis principais no Cluster 1 (Homem, jovem, pardo, solteiro) e Cluster 2 (Homem, mais velho, casado, branco, baixa escolaridade).
- **Meios Letais e Jovens:** A alta associação (66% superior ao acaso) entre a faixa etária 15-29 anos e óbito por arma de fogo indica que o controle de acesso a meios letais é uma estratégia crítica de prevenção.

As estratégias de prevenção devem ser duplamente focadas:

- **Ações de Saúde Mental Jovem (Cluster 1):** Suporte psicológico e campanhas específicas voltadas para o público masculino jovem e solteiro.
- **Ações de Suporte ao Homem Adulto (Cluster 2):** Programas de apoio à saúde mental voltados para homens mais velhos e casados, um grupo de alto risco frequentemente negligenciado.

5 Referências

ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE. Pandemia de COVID-19 desencadeia aumento de 25% na prevalência de ansiedade e depressão. OPAS/OMS, 2 mar. 2022. Disponível em: <https://www.paho.org/pt/noticias/2-3-2022-pandemia-covid-19-desenca>. Acesso em: 20 out. 2025.

BRASIL. Ministério da Saúde. Boletim Epidemiológico, v. 55, n. 4, 2024. Disponível em: <https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/epidemiologicos/edicoes/2024/boletim-epidemiologico-volume-55-no-04.pdf>. Acesso em: 20 out. 2025.

FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA. Anuário Brasileiro de Segurança Pública 2023. 17. ed. São Paulo, 2023. Disponível em: <https://forumseguranca.org.br/wp-content/uploads/2023/07/anuario-2023.pdf>. Acesso em: 20 out. 2025.

BRASIL. Ministério da Saúde. DATASUS. Portal da Saúde. Disponível em: <https://datasus.saude.gov.br/>. Acesso em: 20 out. 2025.

BRASIL. Ministério da Saúde. DATASUS. Sistema de Informações sobre Mortalidade - CID-10. Disponível em: <https://datasus.saude.gov.br/mortalidade-desde-1996-pela-cid-10/>. Acesso em: 20 out. 2025.

BRASIL. Ministério da Saúde. DATASUS. Classificação Internacional de Doenças - CID-10. Disponível em: <http://www2.datasus.gov.br/cid10/V2008/cid10.htm>. Acesso em: 20 out. 2025.

KLANN, V. BrSuicides-dataset: repository made to facilitate time series analysis on Brazil's daily suicide cases, provided by DATASUS. 2025. Repositório de dados e código-fonte. Disponível em: <https://github.com/vklannusp/BrSuicides-dataset>. Acesso em: 4 dez. 2025.

KLANN, V. W.; SOUSA, E. P. M. de. BrSuicides: Conjunto de dados para análise de casos diários de suicídio no Brasil. In: DATASET SHOWCASE WORKSHOP (DSW), 6., 2024, Florianópolis. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2024. p. 12-22. DOI: 10.5753/dsw.2024.244020. Disponível em: <https://sol.sbc.org.br/index.php/dsw/article/view/30611>. Acesso em: 4 dez. 2025.