

SHEILA GENESINE DADA

Data mining espaço-temporal dos suicídios nas cidades brasileiras

São Paulo

2023

SHEILA GENESINE DADA

Data mining espaço-temporal dos suicídios nas cidades brasileiras

Versão Corrigida

Dissertação apresentada à Escola
Politécnica da Universidade de São Paulo
para obtenção do título de Mestre em
Ciências

São Paulo

2023

SHEILA GENESINE DADA

Data mining espaço-temporal dos suicídios nas cidades brasileiras

Versão Corrigida

Dissertação apresentada à Escola
Politécnica da Universidade de São Paulo
para obtenção do título de Mestre em
Ciências

Área de Concentração: Engenharia de
Computação

Orientador: Prof. Dr. Antonio Mauro
Saraiva

São Paulo

2023

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 30 de novembro de 2023

Assinatura do autor: Sheila G. Dada

Assinatura do orientador: Assunção

Catálogo-na-publicação

Dada, Sheila Genesine

Data mining espaço-temporal dos suicídios nas cidades brasileiras / S. G.

Dada -- versão corr. -- São Paulo, 2023.

136 p.

Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Computação e Sistemas Digitais.

1.Suicídio (Brasil) 2.Análise de séries temporais 3.Agrupamento espaço temporal I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Computação e Sistemas Digitais II.t.

***À minha família,
pelo imenso apoio.***

AGRADECIMENTOS

Agradeço aos meus pais Antônio e Catarina pelo exemplo, apoio e palavras sempre acolhedoras.

À minha irmã Thaís, por ser minha inspiração e bússola.

Ao Marco, pelo amor, apoio, parceria e cuidado.

Ao Professor Antonio Saraiva, por levar o ensino à excelência e ter me dado a honra de sua orientação. Pela ajuda e preocupação em todos os momentos e pelas palavras sempre doces e motivadoras.

Ao grande mentor, líder e amigo, Etienne Cartolano, pelo apoio e motivação nesta jornada.

Aos Professores Ligia Vizeu, Paulo Saldiva e Marco Scanavino, pelos grandes conselhos e orientação no grupo *Suicities*.

À Monica Rodrigues, pesquisadora do grupo *Suicities*, pela parceria e troca de conhecimento.

Aos Professores Fabio Cozman e Anna Reali, pelo apoio e exemplo.

À Mariza Leone, colega de departamento, sempre doce e prestativa.

Aos Professores Juliana Saragiotto e Professor Airton Stein, pelas contribuições valiosas ao meu mestrado, enriquecendo e acolhendo minha jornada acadêmica.

Aos amigos do Laboratório de Automação Agrícola, em especial ao Fernando Xavier, pelo companheirismo e ajuda.

RESUMO

DADA, S. G. **Data mining espaço-temporal dos suicídios nas cidades brasileiras.** 127p. 2023. Dissertação (Mestrado) – Escola Politécnica. Universidade de São Paulo, 2023

No mundo, aproximadamente 800.000 pessoas se suicidam a cada ano e, somente nos últimos 45 anos, a taxa de suicídio cresceu cerca de 60%. No Brasil, a taxa de suicídio é mais baixa que a mundial, mas o país apresenta um aumento da taxa entre os adultos homens (25-44 anos). Para desenvolver planos de prevenção ao suicídio, com priorização dos recursos nos grupos mais vulneráveis, é importante identificar quais fatores podem influenciar no aumento das taxas de suicídio e quais subpopulações têm um risco maior de crescimento desta taxa. Por isso, é proposta uma metodologia de análise baseada em ciência de dados que possibilite investigações de padrões espaço-temporais e demográficos na taxa de suicídio das cidades do Brasil, para o pesquisador e/ou gestor público do município poder entender melhor sobre o grupo mais atingido da região de interesse. Para viabilizar o estudo, bases de dados do Censo e do Sistema de Informação sobre Mortalidade (SIM), de 1996 a 2019, foram analisadas. Para o estudo espaço-temporal foi realizado o agrupamento espaço-temporal pelo método de Kulldorff e a análise das tendências temporais pela regressão de Joinpoint. A correlação entre o suicídio e taxa de urbanização dos municípios foi feita pelo teste qui-quadrado. Como resultados, observou-se que em todas as regiões do Brasil há ao menos um agrupamento de alto risco de suicídio, com destaque para a região sul que é um agrupamento que contempla quase toda a região. Os agrupamentos de baixo risco estão nas regiões litorâneas, próximos às grandes capitais e à região metropolitana de São Paulo. Historicamente, os municípios de Três Lagoas (Mato Grosso do Sul), Santana de Parnaíba e Avaré, ambos em São Paulo, apresentam maiores tendências de crescimento da taxa. Em contrapartida, as cidades de São Paulo (Capital), Itapeverica da Serra (São Paulo) e Caarapó (Mato Grosso do Sul), apresentam as maiores tendências de diminuição da taxa. Não há indícios de correlação forte da taxa de urbanização de um município com a taxa de suicídio. Espera-se, como sequência deste trabalho investigativo, a realização de estudos empíricos in loco e a formulação de políticas públicas direcionadas para cada cidade de acordo com a sua necessidade. Ademais, almeja-se simplificar o acesso a dados e facilitar a compreensão sobre suicídio para as Secretarias Municipais e Estaduais de Saúde por meio de um painel Tableau de livre acesso atualizado anualmente.

Palavras-Chave: Suicídio. Município. Brasil. Agrupamento Espaço-temporal. Série Temporal. Evento raro.

ABSTRACT

DADA, S. G. **Spatio-temporal data mining of suicide in Brazilian cities**. 127p. 2023. Dissertation (Master Science). Escola Politécnica. Universidade de São Paulo, São Paulo, 2023

Globally, an estimated 800,000 individuals succumb to suicide annually, with the suicide rate witnessing an approximate 60% surge in the past 45 years. While Brazil's suicide rate remains below the global average, a pronounced increment is observed among adult males aged between 25-44 years. For efficacious suicide prevention strategies, it is imperative to delineate the determinants influencing suicide rates and discern subpopulations at heightened risk. This study seeks to introduce a methodological framework utilizing data science to probe into the spatio-temporal and demographic intricacies of suicide rates across Brazilian municipalities. Such an approach aims to offer municipal researchers and/or public administrators enhanced comprehension of the demographic groups most afflicted. The study leveraged datasets from the Census and the Mortality Information System (SIM) spanning 1996 to 2019. For the spatio-temporal analysis, Kulldorff's spatio-temporal clustering and Joinpoint regression were employed for trend analysis. The correlation between municipal urbanization and suicide rates was assessed via the chi-square test. Every Brazilian region exhibited at least one high-risk suicide cluster, with the southern region being notably pervasive. Contrastingly, low-risk clusters were predominantly found proximate to major coastal capitals and the metropolitan vicinity of São Paulo. Over time, municipalities like Três Lagoas (Mato Grosso do Sul), Santana de Parnaíba, and Avaré (both in São Paulo) manifested escalating rate trajectories. In contrast, São Paulo (Capital), Itapequerica da Serra (São Paulo), and Caarapó (Mato Grosso do Sul) indicated pronounced declining trends. The study found no substantial correlation between a municipality's urbanization rate and its suicide prevalence. As a culmination of this research endeavor, it is anticipated that empirical field studies will be initiated and tailored public policies formulated per municipal requirements. An ancillary objective entails streamlining data accessibility and augmenting suicide-related comprehension for Municipal and State Health Departments via an annually updated, publicly available Tableau dashboard.

Keywords: Suicide. Brazil. Spatio-Temporal Clustering. Time Series. Rare Event.

LISTA DE FIGURAS

Figura 1-1 – Histórico da taxa de suicídio ajustado pela idade (a cada 100.000).	17
Figura 1-2 – Metodologia da pesquisa adotada	21
Figura 2-1 - Fatores de risco do suicídio por estágio de vida.....	24
Figura 2-2 - Taxa de suicídio por 100.000 aberta por gênero em 2016. À esquerda o mapa coroplético da taxa de suicídio de mulheres e à direita, o de homens.	25
Figura 2-3 – Informações no Tabnet, fonte de dados e destaque para fontes utilizadas no estudo.....	28
Figura 2-4 – Questionário do Censo para mapeamento de sexo e idade.	31
Figura 2-5 – Resumo ilustrativo do algoritmo proposto por Kulldorff.....	38
Figura 2-6 – Exemplo da saída do <i>software</i> Joinpoint.....	48
Figura 2-7 – Passo a passo resumido da regressão de Joinpoint.....	50
Figura 3-1 – Desenho da arquitetura da solução	58
Figura 3-2 – Processo de captura das informações de município.....	59
Figura 3-3 – Processo de captura das informações de população.....	61
Figura 3-4 – Processo de captura das informações de urbanização.....	63
Figura 3-5 – Processo de captura das informações de suicídio	65
Figura 3-6 – Modelo lógico relacional das bases necessárias para o estudo	66
Figura 3-7 – Número de <i>clusters</i> ideal igual a 2 analisando o <i>elbow</i>	68
Figura 3-8 – Diferença entre Homocedasticidade (à esquerda, com a variância constante ao longo do tempo) e Heterocedasticidade (à direita, com a variância aumentando ao longo do tempo).....	73
Figura 3-9 – Histórico da taxa de suicídio no município Silves (código município 130400).	74
Figura 3-10 – Histórico da taxa de suicídio no município Tijucas do Sul (código do município 412760).....	75
Figura 3-11 – Histórico da taxa de suicídio no município Cássia dos Coqueiros (código do município 351090).	76
Figura 3-12 – Histórico da taxa de suicídio no município Santa Inês (código do município 4123600).....	76
Figura 3-13- Histórico da taxa de suicídio no município Água Santana (código do município 430005).....	77

Figura 3-14 - Relação entre o erro quadrático médio (MSE) e a amplitude de APC significativamente diferente de zero. Elipse vermelha aponta ponto de inflexão considerado para as análises.....	78
Figura 4-1 – Taxa de suicídio no Brasil calculada neste estudo e o publicado pela OMS.....	80
Figura 4-2 – Taxa de suicídio no Brasil por gênero feminino/masculino.	81
Figura 4-3 – Taxa de suicídio em cada faixa de idade por 100.00 habitantes.	82
Figura 4-4 – Taxa de suicídio por 100.000 habitantes por gênero masculino/feminino e faixa de idade.	83
Figura 4-5 – Somatória da distância quadrada de cada ponto ao seu centróide a cada número de clusters determinado. Não se enxerga nessa figura o <i>Elbow</i>	86
Figura 4-6 – Ilustração dos agrupamentos das séries temporais obtidos da taxa de suicídio entre os municípios. Cada gráfico é um agrupamento, o eixo x está o tempo (anos) e o eixo y é a taxa de suicídio normalizada. As linhas pretas são as séries temporais de cada município e a linha vermelha é a série temporal do centróide do agrupamento.....	87
Figura 4-7 – Mapa coroplético da taxa de suicídio padronizada por gênero masculino/feminino e idade no Brasil em 2019 por município.....	89
Figura 4-8 – Mapa Coroplético da taxa padronizada de suicídio por gênero masculino/feminino e idade nos municípios brasileiros em 1996, 2010 e 2019.	90
Figura 4-9 - Região Norte: comparação de 2010 esquerda e 2019 direita. Quanto mais vermelha a cor, maior a taxa de suicídio, quanto mais azul, menor a taxa.....	90
Figura 4-10 - Taxa de suicídio nos municípios da região Nordeste em 2010 (esquerda) e 2019 (direita). Quanto mais vermelha a cor, maior a taxa de suicídio, quanto mais azul, menor a taxa.	91
Figura 4-11 – Taxa de suicídio nos municípios da região Sudeste em 2010 (esquerda) e 2019 (direita). Quanto mais vermelha a cor, maior a taxa de suicídio, quanto mais azul, menor a taxa.	92
Figura 4-12 – Taxa de suicídio nos municípios da região Sul em 2010 (esquerda) e 2019 (direita). Quanto mais vermelha a cor, maior a taxa de suicídio, quanto mais azul, menor a taxa.	92
Figura 4-13 – Taxa de suicídio nos municípios da região Centro-Oeste em 2010 (esquerda) e 2019 (direita). Quanto mais vermelha a cor, maior a taxa de suicídio, quanto mais azul, menor a taxa.	93
Figura 4-14 – Agrupamentos de alto risco, em vermelho, e de baixo risco, em azul.....	95
Figura 4-15 – Agrupamento de risco elevado na região Norte na divisa do Brasil com Peru e Colômbia.....	96

Figura 4-16 – Agrupamento de risco elevado na região Nordeste.	97
Figura 4-17 – Agrupamento de alto risco na região Centro-Oeste e Sudeste.	98
Figura 4-18 – Agrupamento de alto risco na região sul de Minas Gerais.	99
Figura 4-19 – Agrupamento de alto risco na região sul do Brasil., a conclusão não se aplica aos países vizinhos.	100
Figura 4-20 – Agrupamento de baixo risco no Pará e Maranhão.	101
Figura 4-21 – Agrupamento de baixo risco no Maranhão	102
Figura 4-22 – Agrupamento de baixo risco em Fortaleza e arredores.	103
Figura 4-23 – Agrupamento de baixo risco no litoral do nordeste brasileiro e arredores.	104
Figura 4-24 – Agrupamento de baixo risco no Rio de Janeiro.	105
Figura 4-25 – Agrupamento de baixo risco na região metropolitana de São Paulo.	106
Figura 4-26 – Agrupamentos espaço-temporais de alto risco (vermelho) e baixo risco (azul) no Brasil de 1997 a 2019. Cada círculo pequeno representa um município.	107
Figura 4-27 – Três Lagoas, município com maior APC com 0 <i>joinpoint</i>	110
Figura 4-28 – Avaré, segundo município com maior APC com 0 <i>joinpoint</i>	111
Figura 4-29 – Caarapó, município com menor APC (APC mais negativo) com 0 <i>joinpoint</i>	111
Figura 4-30 – São Francisco do Sul, município com o segundo menor APC com 0 <i>joinpoint</i>	112
Figura 4-31 – Município de Campo Grande (código município 500270), apresenta um aumento significativo na tendência de aumento de suicídio. Primeiro maior município com APC com 1 <i>joinpoint</i>	113
Figura 4-32 – Município de Garça (código município 351670), segundo maior PAC com 1 <i>joinpoint</i>	113
Figura 4-33 – Município de Beberibe. Município com a maior tendência de decréscimo no APC com 1 <i>joinpoint</i>	114
Figura 4-34 – Município de Mossoró. Município com a segunda maior queda de APC com 1 <i>joinpoint</i>	115
Figura 4-35 – Município Cascavel, com maior APC com 2 <i>joinpoints</i>	116
Figura 4-36 – Município São Paulo com menor APC com 2 <i>joinpoints</i>	116
Figura 4-37 – Município Santana de Parnaíba com tendência de crescimento com 3 <i>joinpoints</i>	117

Figura 4-38 – Município Itapeceira da Serra com tendência de queda com 3 <i>joinpoints</i>	118
Figura 4-39 – Proporções entre os tipos de urbanização e os agrupamentos de alto e baixo risco.	119

LISTA DE TABELAS

Tabela 2.1 - Fatores de risco do suicídio e exemplos.	27
Tabela 3.1 – Etapas do estudo e proposta de abordagem técnica de cada etapa.	55
Tabela 4.1 – Taxa padronizada do suicídio e comparação do aumento das taxas de 2019 e 1996 nas 25 primeiras cidades mais populosas do Brasil.	84
Tabela 4.2 – Resumo da quantidade de municípios que apresentam aumento, diminuição ou estabilidade na taxa de suicídio ajustada por idade e gênero masculino/feminino.....	110

LISTA DE SÍMBOLOS

Z	Agrupamento espacial analisado (zona)
G	Espaço geográfico total
p	Probabilidade de um evento ocorrer no espaço Z
q	Probabilidade de um evento ocorrer no espaço $G-Z$
μ	Unidade de medida total. Exemplo: $\mu(G)$ = tamanho da população no espaço geográfico G (população total)
η	Número de eventos. Exemplo: η_G = número de eventos que ocorreram no espaço G (suicídios)
\mathcal{X}_i	Unidade de espaço. Exemplo: cada município
I	Função identidade

SUMÁRIO

1	INTRODUÇÃO E JUSTIFICATIVA	17
1.1	OBJETIVOS	20
1.2	METODOLOGIA DA PESQUISA	21
1.3	ESTRUTURAÇÃO DO TEXTO	22
2	REVISÃO BIBLIOGRÁFICA.....	23
2.1	ESTUDO SOBRE O SUICÍDIO	23
2.2	VARIÁVEIS ANALÍTICAS	27
2.2.1	DATASUS	27
2.2.2	CENSO DEMOGRÁFICO	29
2.3	ESTUDO ESPAÇO-TEMPORAL	32
2.3.1	Visão geral	32
2.3.1.1	Pontos	34
2.3.1.2	Séries temporais.....	39
2.3.2	Estudo espaço-temporal do suicídio.....	40
2.3.2.1	Definição dos dados	41
2.3.2.2	Método do cálculo da taxa de suicídio.....	43
2.3.2.3	Análise Descritiva	44
2.3.2.4	Análise Espacial	45
2.3.2.5	Análise Temporal.....	46
2.3.2.6	Análise Multivariada	51
3	MATERIAL E MÉTODOS	53
3.1	DADOS	58
3.2	CÁLCULO DA TAXA DE SUICÍDIO	66
3.3	VALIDAÇÃO DOS DADOS E ANÁLISE DESCRITIVA.....	67
3.4	ANÁLISE ESPACIAL	69
3.5	ANÁLISE TEMPORAL	72

3.6	ANÁLISE MULTIVARIADA.....	78
4	RESULTADOS	80
4.1	CÁLCULO DA TAXA DE SUICÍDIO E ANÁLISE DESCRITIVA	80
4.2	ANÁLISE ESPACIAL	94
4.3	ANÁLISE TEMPORAL	109
4.4	ANÁLISE MULTIVARIADA.....	118
5	CONCLUSÕES E PESQUISAS FUTURAS.....	120
	REFERÊNCIAS BIBLIOGRÁFICAS	127
	GLOSSÁRIO.....	136

1 INTRODUÇÃO E JUSTIFICATIVA

No mundo, aproximadamente 800.000 pessoas se suicidam a cada ano (WHO, 2020a) e, somente nos últimos 45 anos, a taxa de suicídio cresceu cerca de 60% (BERTOLOTE; FLEISCHMANN, 2005). A Organização Mundial da Saúde (OMS) (*The World Health Organization - WHO*) considera o suicídio uma prioridade de Saúde Pública (WHO, 2020a).

A taxa internacional de suicídio flutua entre 10 e 15 a cada 100.000 pessoas. No Brasil, a taxa flutua entre 6 e 7 mortes por 100.000 pessoas. Apesar de a taxa brasileira ser considerada baixa, ela está crescendo a cada ano entre os adultos mais jovens, principalmente homens (KRUG et al., 2002). A Figura 1-1 exemplifica o crescimento da taxa no Brasil. Além disso, há uma heterogeneidade no país em relação à taxa de suicídio, sendo que 18% dos municípios brasileiros apresentaram em 2019 taxa de suicídio maior que do que a média global.

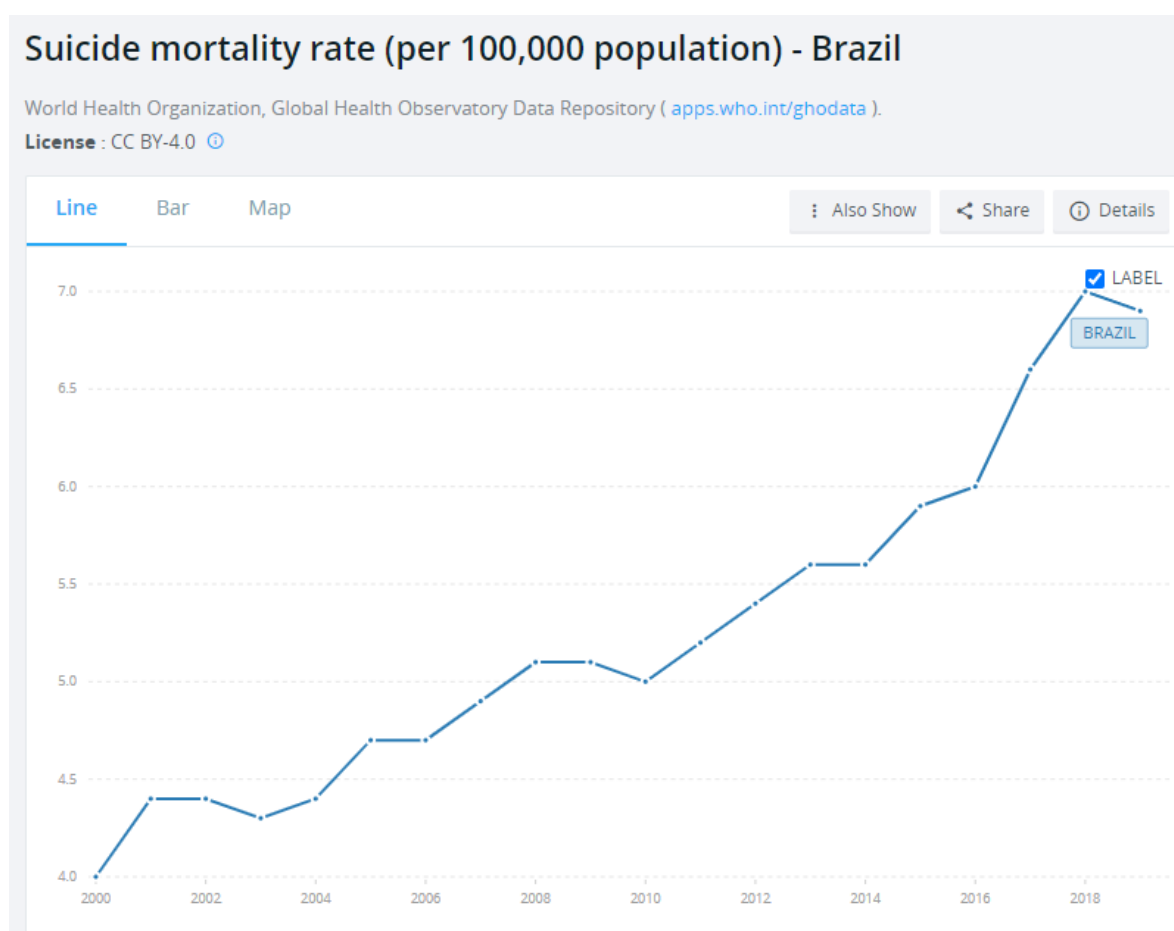


Figura 1-1 – Histórico da taxa de suicídio ajustado pela idade (a cada 100.000).
Fonte: THE WORLD BANK (2023)

Os fatores de risco para o suicídio podem ser classificados como: **(1)** riscos relacionados ao nível individual (ex.: transtornos mentais), **(2)** riscos relacionados às comunidades e aos relacionamentos (ex.: estresse entre migrantes), **(3)** riscos sociodemográficos (ex.: homens, divorciados) e **(4)** riscos ambientais (ex.: violência urbana, congestionamentos, poluição sonora e visual e mudanças meteorológicas diárias).

Os fatores individuais, com destaque para os transtornos psicológicos, demonstram uma influência substancial nas taxas de suicídio. O suicídio, à semelhança de outras condições psiquiátricas, exibe uma tendência a manifestar-se em contextos familiares. Distúrbios como depressão, transtorno bipolar, transtornos do espectro da esquizofrenia, transtornos relacionados ao uso de substâncias, epilepsia e lesão cerebral traumática têm sido associados a um aumento de aproximadamente três vezes na probabilidade de suicídio consumado. É reconhecido que o suicídio não pode ser reduzido unicamente à presença de transtornos psicológicos (predisposição). É compreendido que a coexistência da predisposição ao suicídio, em conjunto com fatores precipitantes (com riscos 2 a 4, conforme mencionado no parágrafo anterior) e a disponibilidade facilitada de métodos letais, podem potencializar a concretização de atos suicidas (FAZEL; RUNESON, 2020).

Existe uma previsão de que os fatores de precipitação possam se agravar nos anos subsequentes. Há projeções que indicam um possível aumento das taxas de suicídio devido às mudanças climáticas, derivando tanto de variáveis naturais como antropogênicas. A alteração climática contribui para um aumento na ocorrência de eventos de desastres naturais, cuja relação com o aumento de doenças mentais tem sido documentada (HAYES et al., 2018). Além disso, as mudanças climáticas podem resultar em escassez de recursos hídricos e alimentares, bem como no agravamento de enfermidades, o que pode acarretar maior estresse na população. Indiretamente, esses impactos ambientais também estão associados à sensação de isolamento e impotência entre os indivíduos, dada a percepção de falta de ações eficazes tanto no âmbito individual como coletivo para mitigar a situação atual (HAYES et al., 2018).

O período atual, conhecido como Antropoceno, é caracterizado pelo ser humano como o principal agente das mudanças observadas no planeta. A partir da Segunda Guerra Mundial, observa-se um aumento expressivo nos padrões de consumo, crescimento exponencial da população e aceleração dos impactos humanos sobre o meio ambiente e os sistemas naturais. As mudanças climáticas e a pandemia da COVID-19 são exemplos ilustrativos de como os sistemas naturais são afetados e, por sua vez, impactam a saúde e o bem-estar da população, incluindo a Saúde Mental. Dado que o suicídio não é meramente um ato individual, desvinculado de influências externas, a crise ambiental em curso pode contribuir para um aumento na incidência desses eventos (WHITMEE et al., 2015).

Um campo de pesquisa transdisciplinar denominado "Saúde Planetária" surgiu com uma orientação voltada para a proposição de soluções a análises de enfrentamento dos impactos das perturbações causadas pela atividade humana nos sistemas naturais do planeta e seu impacto na saúde humana e em toda a biodiversidade. A Saúde Planetária ressalta a urgência de abordar esses problemas, como o suicídio, que são de natureza complexa, transcende disciplinas e têm alcance global (WHITMEE et al., 2015).

Como mencionado anteriormente, existe a possibilidade de um aumento futuro nas taxas de suicídio. Para contrabalançar esse cenário previsto, o Brasil e outros países adotaram um compromisso orientado pelo Plano de Ação de Doença Mental da Organização Mundial de Saúde, o qual estabelece a meta de reduzir em 33% a taxa de suicídio até o ano de 2030, tomando como referência a taxa de base do ano 2012/2013 (WHO, 2021).

Para elaborar estratégias de prevenção do suicídio, com a alocação de recursos priorizados para grupos de maior vulnerabilidade, torna-se pertinente a identificação dos seguintes aspectos: (1) os fatores que possam estar associados ao aumento das taxas de suicídio, (2) as subpopulações que apresentam um risco elevado de aumento dessas taxas (JAEN-VARAS et al., 2019), e (3) a análise da variação temporal e espacial do comportamento do suicídio nas áreas urbanas do Brasil. Além disso, é recomendável identificar grupos que apresentam uma menor propensão ao suicídio, uma vez que podem oferecer *insights* sobre práticas eficazes de prevenção (RODRIGUES et al., 2019). A disponibilização de uma ferramenta de

fácil acesso tanto para pesquisadores quanto para gestores públicos é crucial para esse propósito.

1.1 OBJETIVOS

O objetivo geral desta pesquisa é propor uma metodologia de análise de ciência de dados que possibilitem investigações de padrões espaço-temporais e demográficos na taxa de suicídio das cidades do Brasil.

Os objetivos específicos são:

1. identificar os agrupamentos espaço-temporais de maior risco de suicídio;
2. identificar as cidades com maiores crescimentos ou quedas históricas da taxa;
3. verificar os fatores de risco individuais (dados demográficos e gênero e coletivos (percentual de urbanização) de suicídio;
4. desenvolver uma ferramenta prática e objetiva de apoio para os gestores públicos e pesquisadores no manejo do suicídio.

O estudo é direcionado pelas seguintes perguntas:

5. Quais agrupamentos espaciais podem estar relacionados com os dados sobre suicídio? Esses agrupamentos mudam ao longo do tempo? Há algum agrupamento de maior relevância? Há agrupamentos de baixo risco para destacar?
6. Historicamente, quais são as cidades com maiores crescimentos da taxa e as com maiores quedas?
7. Quais municípios e quais grupos de gênero masculino/feminino e idade contribuem para o aumento da taxa de suicídio no Brasil?
8. O percentual de urbanização contribui para explicar as variações das taxas de suicídio?

Para responder essas perguntas, são utilizados os dados do Censo demográfico (IBGE, 2020b) e do Departamento de Informática do Sistema Único de Saúde – DATASUS (DATASUS, 2019b) de 1996 a 2019.

1.2 METODOLOGIA DA PESQUISA

A Figura 1-2 ilustra, esquematicamente, a metodologia da pesquisa adotada.

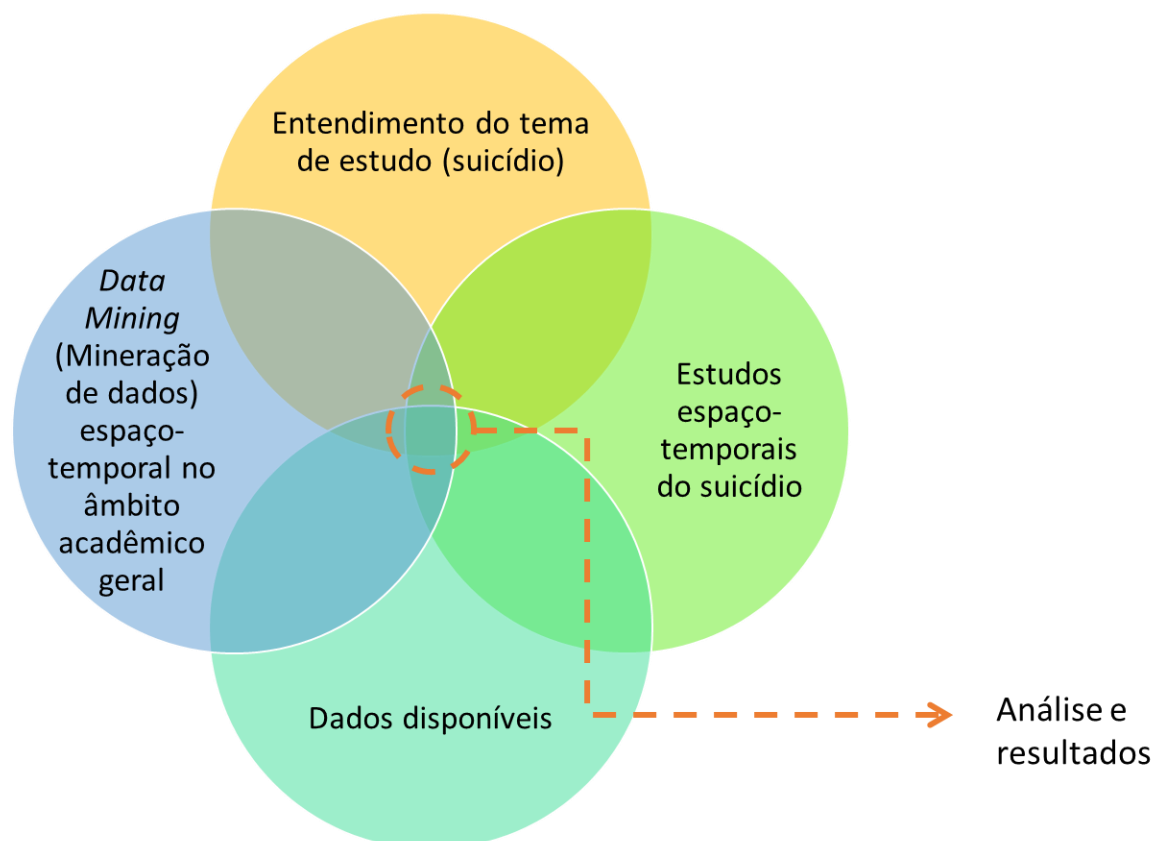


Figura 1-2 – Metodologia da pesquisa adotada

A metodologia consiste em quatro elementos principais que são insumos para as análises e resultados:

1. Entendimento do tema de estudo (suicídio): entendimento teórico de quais são as causas do suicídio para se poder propor a metodologia e ferramenta posterior que melhor se adequem à questão.
2. Data Mining espaço-temporal no âmbito acadêmico geral: aprofundamento dos métodos e ferramentas de ciência de dados utilizados na literatura para se resolverem problemas espaço-temporais em diversas áreas de aplicação.
3. Estudo espaço-temporais do suicídio: levantamento de estudos espaço-temporais já realizados na academia para se entender como o

suicídio se comporta no espaço e tempo em diversas regiões do mundo.

4. Dados disponíveis: mapeamento dos dados disponíveis no Brasil para se alcançar o objetivo desejado.

Considerando esses quatro elementos, desenvolve-se a captura, transformação e agregação dos dados, para posterior análise, desenvolvimento da ferramenta e exposição dos resultados.

1.3 ESTRUTURAÇÃO DO TEXTO

As demais partes do texto estão assim estruturadas:

- Capítulo 2 – Revisão Bibliográfica: engloba aspectos gerais sobre (1) o embasamento teórico do suicídio; (2) o embasamento teórico dos métodos estatísticos que foram utilizados em estudos de geografia da saúde em revistas acadêmicas de saúde pública e (3) o estudo de novas propostas de metodologias estatísticas baseadas em revistas acadêmicas de Computação e Estatística;
- Capítulo 3 – Material e Métodos: apresenta o conjunto de dados, ferramentas e abordagem estatística e computacional do estudo;
- Capítulo 4 – Resultados: apresenta o resultado da análise espaço temporal;
- Capítulo 5 – Conclusões e pesquisas futuras: apresenta as conclusões e aponta oportunidades de novas pesquisas;

2 REVISÃO BIBLIOGRÁFICA

Os subitens a seguir apresentam uma revisão bibliográfica sobre os quatro principais insumos citados em 1.2 METODOLOGIA DA PESQUISA para se propor uma metodologia de ciência de dados no estudo do suicídio espaço-temporal.

2.1 ESTUDO SOBRE O SUICÍDIO

Durkheim ([1897] 2000), pesquisador clássico da Sociologia, defendeu que o suicídio é um fenômeno coletivo, social. Ele fundamentou que "existe para cada grupo social uma tendência específica para o suicídio que nem a constituição orgânica-psíquica dos indivíduos, nem a natureza do ambiente natural explicam, resultando disso, por eliminação, que essa tendência deva depender de causas sociais e constituir por si mesma um fenômeno coletivo."

O autor buscou formas de encontrar as causas do suicídio como um fenômeno social. Ele utilizou informações de mortalidade, riqueza da região, gênero masculino/feminino, estado civil, religião e educação para encontrar correlações.

Além do estudo sistêmico, Durkheim ([1897] 2000) definiu quatro tipos de suicídio, são eles:

- a) **Altruísta:** indivíduo integra-se em excesso à sociedade, ao ponto em que há algo exterior a ele que vale mais que sua vida. Por exemplo: homem bomba.
- b) **Egoísta:** indivíduo não se identifica com a sociedade e não encontra razões para continuar vivendo. Por exemplo: pessoas depressivas.
- c) **Anômico:** indivíduo sente a quebra da ordem social e recorre ao suicídio para mudar seu estado social. Exemplo: crise financeira.
- d) **Fatalista:** indivíduo se sente preso à ordem social e recorre ao suicídio para escapar de um possível cenário social imposto. Exemplo: escravidão.

Fazel e Runeson (2020) recentemente publicaram uma revisão bibliográfica sobre suicídio. Eles defendem que os fatores de risco para o suicídio são divididos entre o nível individual e o populacional (ambiental).

Apesar dos avanços dos estudos desde 1897, os fatores sociais/populacionais ainda se mostram importantes para entender melhor o suicídio.

Fazel e Runeson (2020) defendem a análise do curso de vida do indivíduo para compreender os fatores de risco do suicídio. Essa abordagem defende que o fator de risco varia de acordo com os diferentes estágios de vida e o suicídio é o resultado cumulativo de diversos fatores de risco. A Figura 2-1 apresenta esse movimento, em que cada seta é um fator de risco. Quanto mais forte a cor da seta na figura, maior o impacto do fator de risco no determinado estágio da vida.

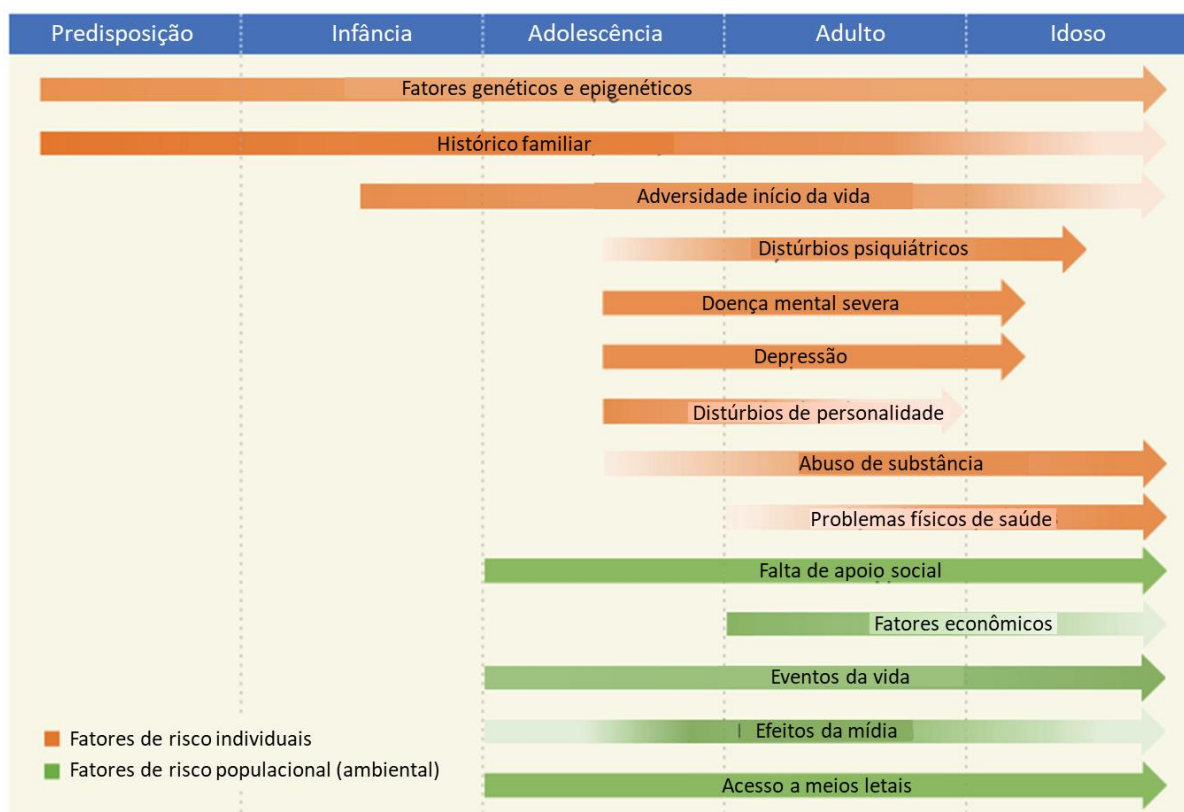


Figura 2-1 - Fatores de risco do suicídio por estágio de vida.
Fonte: FAZEL; RUNESON (2020)

Além da variação da taxa de suicídio pelos estágios da vida (faixa etária) por conta de diferentes fatores de risco, observa-se também que a taxa de mortalidade é maior para homens do que para mulheres (WHO, 2016), conforme Figura 2-2.

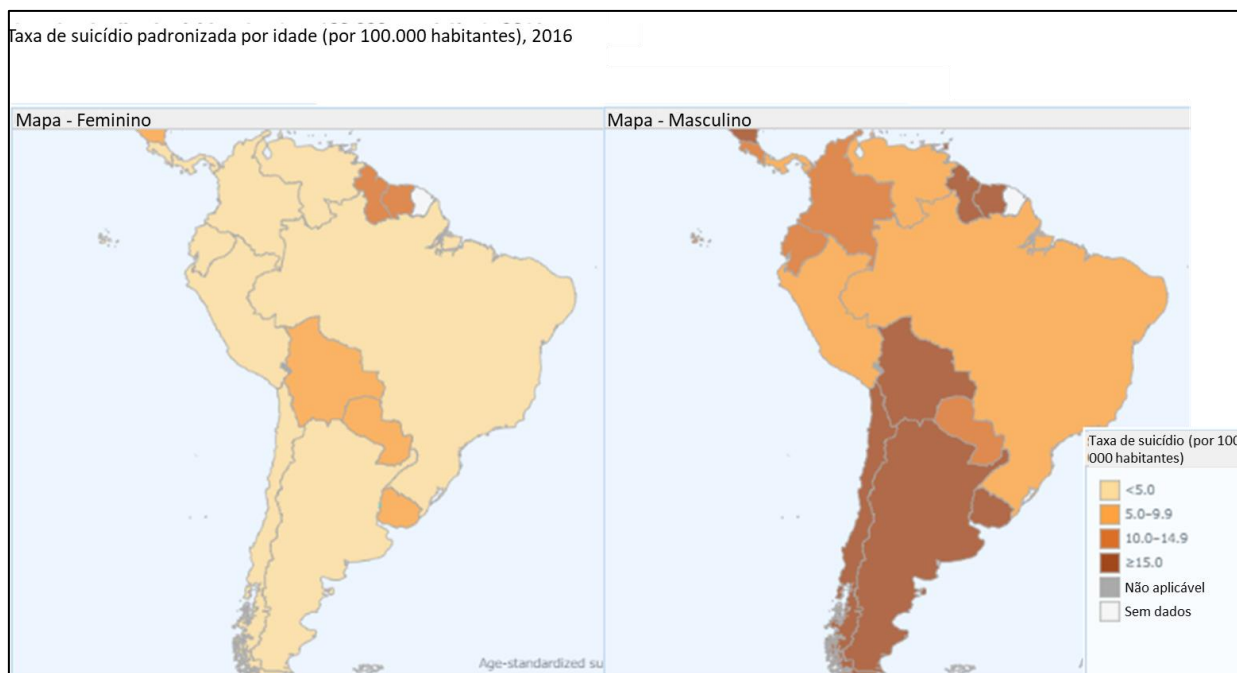


Figura 2-2 - Taxa de suicídio por 100.000 aberta por gênero em 2016. À esquerda o mapa coroplético da taxa de suicídio de mulheres e à direita, o de homens.
Fonte: WHO (2016)

Conforme previamente postulado por Durkheim ([1897] 2000), a explicação para o ato de suicídio não se restringe unicamente à constituição orgânico-psíquica dos indivíduos. Portanto, argumenta-se que há fatores de predisposição que interagem com fatores precipitantes, desencadeando alterações psicológicas. Essas alterações psicológicas conduzem ao isolamento social, que, quando combinado com o acesso a métodos letais, pode culminar no ato suicida (VAN ORDEN et al., 2010).

Dentre os riscos individuais e de predisposição, encontram-se o distúrbio psiquiátrico, o histórico suicida na família, as tentativas de suicídio anteriores, as experiências adversas na infância e a privação socioeconômica (FAZEL; RUNESON, 2020).

Sobre a privação socioeconômica, em países desenvolvidos, foi comprovado que pessoas com menor renda apresentam maior taxa de suicídio (JAEN-VARAS et al., 2019; OKAMOTO et al., 2013). No Brasil, há estudos que apontam que quanto menor a privação social, maior a taxa de suicídio (ALARCÃO et al., 2019; JAEN-VARAS et al., 2019), enquanto outros não indicam correlação entre a taxa de suicídio e fatores socioeconômicos (DANTAS et al., 2018). A hipótese é que a privação social

por um lado gera mais eventos estressantes, mas ela promove menor isolamento social (por exemplo: convivência nas comunidades e maior número de religiosos).

No contexto do suicídio, presume-se que o sujeito manifeste tanto fatores de predisposição quanto fatores de precipitação. Estes últimos referem-se a eventos específicos e estressantes na trajetória de vida de um indivíduo, os quais podem culminar em um episódio de desequilíbrio psicológico, potencializando a concretização do ato suicida. Entre os fatores que aceleram o ato (de precipitação) estão eventos estressantes da vida, como dificuldades no relacionamento (por exemplo, divórcio), imigração, morte de algum ente querido, diagnóstico de câncer e vítimas de agressão e violência, como vítimas de assaltos e prisioneiros. Desastres naturais, mudanças climáticas, filiações religiosas (ou falta delas) e acesso a armas letais também são fatores de precipitação (FAZEL; RUNESON, 2020).

A Tabela 2.1 resume os fatores de risco abordados por Fazel, Runeson (2020).

Fatores de risco	Exemplos
Fatores de predisposição	
Distúrbios psicológicos	Depressão, transtorno bipolar, transtornos do espectro da esquizofrenia, transtornos por uso de substâncias, epilepsia e lesão cerebral traumática
Histórico familiar de comportamento suicida	Auto descritivo
Tentativa anterior de suicídio	Auto descritivo
Experiências adversas de infância	Abuso sexual infantil, perda de parentes
Privação socioeconômica	Menor renda
Fatores de precipitação	
Consumo de álcool e drogas	Auto descritivo
Acesso a armas letais	Acesso pesticidas, construções altas (por exemplo, pontes), medicamentos letais e armas de fogo
Eventos estressantes	Dificuldades no relacionamento (por exemplo, divórcio), imigração, morte de um ente querido, diagnóstico de doenças crônicas e vítimas de agressão e violência.
Efeito da mídia	Divulgação na mídia de suicídio.
Efeitos ambientais	Desastres naturais, mudanças climáticas e picos e vales de temperatura.
Religião	Auto descritivo

Tabela 2.1 - Fatores de risco do suicídio e exemplos.

Após compreender os intrincados fatores que podem levar um indivíduo ao suicídio, é essencial avaliar as variáveis analíticas disponíveis que nos permitem investigar tais fenômenos. Na próxima seção, abordaremos as fontes dessas variáveis e como elas contribuem para a compreensão do suicídio.

2.2 VARIÁVEIS ANALÍTICAS

Os subitens a seguir, apresentam um aprofundamento teórico das fontes de dados utilizadas na pesquisa.

2.2.1 DATASUS

O Departamento de Informática do Sistema Único de Saúde (DATASUS) possui a responsabilidade de prover aos órgãos do Sistema Único de Saúde público

brasileiro, SUS, sistemas de informação e suporte de informática, necessários ao processo de planejamento, operação e controle. O Departamento é responsável pela elaboração e manutenção de sistemas, aplicativos, portais e banco de dados dos órgãos do SUS (DATASUS, 2019a).

Neste estudo, foram utilizadas as informações disponibilizadas pelo Tabnet, aplicativo desenvolvido pelo DATASUS, com o objetivo de gerar as informações das bases de dados do SUS e subsidiar as análises para se pensar em ações de saúde (DATASUS, 2019b).

O Tabnet apresenta dados de morbidade, incapacidade, acesso a serviços, qualidade da atenção, condições de vida, fatores ambientais, assistência à saúde da população, cadastro dos estabelecimentos de saúde, além de dados sobre recursos financeiros e informações demográficas e socioeconômicas (DATASUS, 2019b). A Figura 2-3 ilustra os dados existentes no Tabnet e respectivos sistemas origem desses dados. A seguir são explicadas as fontes de interesse para o estudo.

Origem das Informações

Indicadores de Saúde e Pactuações	IDB / Cadernos de Saúde / Pactuação
Assistência à Saúde	Sistemas Hospitalares / Ambulatoriais / Imunizações / Atenção Básica / Vig. Nutricional
Epidemiológicas e Morbidade	Sistemas de Morbidade / Ag. Avos / Nutrição
Rede Assistencial	Cadastro Nacional Estabelecimento Saúde
Estatísticas Vitais	Sistema de Mortalidade / nascidos Vivos
Demográficas e Socioeconômicas	População/PIB/Saneamento
Inquéritos e Pesquisas	PNAD. VIGITEL/VIVA
Saúde Suplementar (ANS)	Agência Nacional de Saúde Suplementar
Recursos do SUS	
GAP	
Estatísticas de acesso ao TABNET	Estatísticas de acesso






Figura 2-3 – Informações no Tabnet, fonte de dados e destaque para fontes utilizadas no estudo.
Fonte: DATASUS (c2020)

Para se determinar a morte por suicídio, foi utilizada a base do Sistema de Informação sobre Mortalidade (SIM). O sistema é gerido pelo Departamento de Análise de Situação de Saúde, da Secretaria de Vigilância em Saúde, em conjunto com as Secretarias Estaduais e Municipais de Saúde. As Secretarias de Saúde

coletam as Declarações de Óbitos dos cartórios e cadastram essas informações no SIM. Nessas declarações, a causa do óbito é registrada através de um código, que é determinado pela Classificação Internacional de Doenças, versão 10 (CID-10) (TABNET, 2017).

A CID-10 é a classificação padrão do diagnóstico para fins clínicos e investigativos. Ela é determinada pela Organização Mundial de Saúde (OMS) e foi implementada no Brasil em 1996. A CID define o universo de doenças, distúrbios, lesões e outras condições de saúde de maneira hierárquica (WHO, 2020b). A CID é dividida em capítulos, em seguida em grupos, categorias e, por fim, subcategorias. Por exemplo, a CID X600 pertence ao capítulo “XX - Causas externas de morbidade e de mortalidade” (X do X600), ao grupo “Lesões autoprovocadas intencionalmente” (X60 ao X84), à categoria “Autointoxicação por e exposição, intencional, a analgésicos, antipiréticos e antirreumáticos, não-opiáceos” (60 do “X600”) e à subcategoria “residência” (0 do X600). Os CIDs considerados como suicídio neste estudo estão descritos em Material e Métodos na seção 2.3.2.1 Definição dos dados.

Após a análise dos dados de suicídio provenientes do SUS, torna-se essencial considerar os fatores demográficos que podem influenciar tais ocorrências. Na seção subsequente, é apresentado o censo demográfico e sua relevância no contexto do estudo.

2.2.2 CENSO DEMOGRÁFICO

O Censo é uma pesquisa decenal feita em todos os municípios do Brasil. O objetivo da pesquisa é conhecer as condições de vida da população. O Censo mais recente é o de 2022, porém a pesquisa utiliza o Censo de 2010 pois a coleta de dados ocorreu antes da publicação do Censo de 2022. O Censo apresenta as características dos domicílios, identificação étnico-racial, nupcialidade, núcleo familiar, fecundidade, religião ou culto, deficiência, migração interna ou internacional, educação, deslocamento para estudo, trabalho e rendimento, deslocamento para trabalho, e mortalidade da população brasileira. (IBGE, 2020a).

Os microdados e a documentação da base estão disponíveis no IBGE (2020b). Os microdados disponibilizam as respostas já agrupadas em área de ponderação para se preservar o sigilo dos respondentes. Uma área de ponderação é, em geral, uma região dentro dos municípios.

A Figura 2-4 ilustra a etapa do questionário do Censo referente às informações de sexo e idade do cidadão.

2 INFORMAÇÕES SOBRE MORADORES (PARA DOMICÍLIOS PARTICULARES E COLETIVOS)																					
<p>1 2.01 QUANTAS PESSOAS MORAVAM NESTE DOMICÍLIO EM 31 DE JULHO DE 2022?</p> <input type="text"/> <input type="text"/> → Siga para 2.02	<p>2 2.02 QUANTAS CRIANÇAS DE ZERO A NOVE ANOS DE IDADE, INCLUSIVE RECÊM-NASCIDOS, MORAVAM NESTE DOMICÍLIO EM 31 DE JULHO DE 2022?</p> <input type="text"/> <input type="text"/> → Siga para 2.03.1																				
LISTA DE MORADORES EM 31 DE JULHO DE 2022																					
<p>3 2.03.1 NOME DO MORADOR: <input type="text"/></p> <p>2.03.2 SOBRENOME DO MORADOR: <input type="text"/></p>																					
<p>4 2.04 SEXO: <input type="checkbox"/> 1 - MASCULINO <input type="checkbox"/> 2 - FEMININO</p>																					
<p>5 2.05 QUAL É A DATA DE NASCIMENTO?</p>																					
<p>2.05.1 DIA: <input type="text"/> <input type="text"/> 2.05.2 MÊS:</p> <table border="0"> <tr> <td><input type="checkbox"/> 01 - JANEIRO</td> <td><input type="checkbox"/> 05 - MAIO</td> <td><input type="checkbox"/> 09 - SETEMBRO</td> </tr> <tr> <td><input type="checkbox"/> 02 - FEVEREIRO</td> <td><input type="checkbox"/> 06 - JUNHO</td> <td><input type="checkbox"/> 10 - OUTUBRO</td> </tr> <tr> <td><input type="checkbox"/> 03 - MARÇO</td> <td><input type="checkbox"/> 07 - JULHO</td> <td><input type="checkbox"/> 11 - NOVEMBRO</td> </tr> <tr> <td><input type="checkbox"/> 04 - ABRIL</td> <td><input type="checkbox"/> 08 - AGOSTO</td> <td><input type="checkbox"/> 12 - DEZEMBRO</td> </tr> </table> <p>2.05.3 ANO <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 2.05.3A NÃO SEI O MÊS E/OU ANO DE NASCIMENTO <input type="checkbox"/></p> <p>Se assinalar 2.05.3A, siga para 2.05.4 Caso contrário, passe para 2.06</p>		<input type="checkbox"/> 01 - JANEIRO	<input type="checkbox"/> 05 - MAIO	<input type="checkbox"/> 09 - SETEMBRO	<input type="checkbox"/> 02 - FEVEREIRO	<input type="checkbox"/> 06 - JUNHO	<input type="checkbox"/> 10 - OUTUBRO	<input type="checkbox"/> 03 - MARÇO	<input type="checkbox"/> 07 - JULHO	<input type="checkbox"/> 11 - NOVEMBRO	<input type="checkbox"/> 04 - ABRIL	<input type="checkbox"/> 08 - AGOSTO	<input type="checkbox"/> 12 - DEZEMBRO								
<input type="checkbox"/> 01 - JANEIRO	<input type="checkbox"/> 05 - MAIO	<input type="checkbox"/> 09 - SETEMBRO																			
<input type="checkbox"/> 02 - FEVEREIRO	<input type="checkbox"/> 06 - JUNHO	<input type="checkbox"/> 10 - OUTUBRO																			
<input type="checkbox"/> 03 - MARÇO	<input type="checkbox"/> 07 - JULHO	<input type="checkbox"/> 11 - NOVEMBRO																			
<input type="checkbox"/> 04 - ABRIL	<input type="checkbox"/> 08 - AGOSTO	<input type="checkbox"/> 12 - DEZEMBRO																			
<p>2.05.4 IDADE:</p> <p><input type="checkbox"/> 1 - UM ANO OU MAIS <input type="checkbox"/> 2 - MENOS DE UM ANO</p> <p>Se quesito 2.05.4 igual a 1, passe para 2.05.5 Se quesito 2.05.4 igual a 2, passe para 2.05.6</p> <p>2.05.5 IDADE EM ANOS: (1 ANO OU MAIS) <input type="text"/> <input type="text"/> <input type="text"/> → Siga para 2.06</p> <p>2.05.6 IDADE EM MESES: (MENOS DE 1 ANO) <input type="text"/> <input type="text"/> → Siga para 2.06</p>																					
<p>6 2.06 QUAL É A RELAÇÃO DE PARENTESCO OU DE CONVIVÊNCIA COM A PESSOA RESPONSÁVEL PELO DOMICÍLIO?</p> <table border="0"> <tr> <td><input type="checkbox"/> 01 - PESSOA RESPONSÁVEL PELO DOMICÍLIO</td> <td><input type="checkbox"/> 11 - BISNETO(A)</td> </tr> <tr> <td><input type="checkbox"/> 02 - CÔNJUGE OU COMPANHEIRO(A) DE SEXO DIFERENTE</td> <td><input type="checkbox"/> 12 - IRMÃO OU IRMÃ</td> </tr> <tr> <td><input type="checkbox"/> 03 - CÔNJUGE OU COMPANHEIRO(A) DO MESMO SEXO</td> <td><input type="checkbox"/> 13 - AVÔ OU AVÓ</td> </tr> <tr> <td><input type="checkbox"/> 04 - FILHO(A) DO RESPONSÁVEL E DO CÔNJUGE</td> <td><input type="checkbox"/> 14 - OUTRO PARENTE</td> </tr> <tr> <td><input type="checkbox"/> 05 - FILHO(A) SOMENTE DO RESPONSÁVEL</td> <td><input type="checkbox"/> 15 - AGREGADO(A)</td> </tr> <tr> <td><input type="checkbox"/> 06 - ENTEADO(A)</td> <td><input type="checkbox"/> 16 - CONVIVENTE</td> </tr> <tr> <td><input type="checkbox"/> 07 - GENRO OU NORA</td> <td><input type="checkbox"/> 17 - PENSIONISTA</td> </tr> <tr> <td><input type="checkbox"/> 08 - PAI, MÃE, PADRASTO OU MADRASTA</td> <td><input type="checkbox"/> 18 - EMPREGADO(A) DOMÉSTICO(A)</td> </tr> <tr> <td><input type="checkbox"/> 09 - SOGRO(A)</td> <td><input type="checkbox"/> 19 - PARENTE DO(A) EMPREGADO(A) DOMÉSTICO(A)</td> </tr> <tr> <td><input type="checkbox"/> 10 - NETO(A)</td> <td><input type="checkbox"/> 20 - INDIVIDUAL EM DOMICÍLIO COLETIVO</td> </tr> </table> <p style="text-align: right;">→ Encerre o bloco e siga para 3.01</p>		<input type="checkbox"/> 01 - PESSOA RESPONSÁVEL PELO DOMICÍLIO	<input type="checkbox"/> 11 - BISNETO(A)	<input type="checkbox"/> 02 - CÔNJUGE OU COMPANHEIRO(A) DE SEXO DIFERENTE	<input type="checkbox"/> 12 - IRMÃO OU IRMÃ	<input type="checkbox"/> 03 - CÔNJUGE OU COMPANHEIRO(A) DO MESMO SEXO	<input type="checkbox"/> 13 - AVÔ OU AVÓ	<input type="checkbox"/> 04 - FILHO(A) DO RESPONSÁVEL E DO CÔNJUGE	<input type="checkbox"/> 14 - OUTRO PARENTE	<input type="checkbox"/> 05 - FILHO(A) SOMENTE DO RESPONSÁVEL	<input type="checkbox"/> 15 - AGREGADO(A)	<input type="checkbox"/> 06 - ENTEADO(A)	<input type="checkbox"/> 16 - CONVIVENTE	<input type="checkbox"/> 07 - GENRO OU NORA	<input type="checkbox"/> 17 - PENSIONISTA	<input type="checkbox"/> 08 - PAI, MÃE, PADRASTO OU MADRASTA	<input type="checkbox"/> 18 - EMPREGADO(A) DOMÉSTICO(A)	<input type="checkbox"/> 09 - SOGRO(A)	<input type="checkbox"/> 19 - PARENTE DO(A) EMPREGADO(A) DOMÉSTICO(A)	<input type="checkbox"/> 10 - NETO(A)	<input type="checkbox"/> 20 - INDIVIDUAL EM DOMICÍLIO COLETIVO
<input type="checkbox"/> 01 - PESSOA RESPONSÁVEL PELO DOMICÍLIO	<input type="checkbox"/> 11 - BISNETO(A)																				
<input type="checkbox"/> 02 - CÔNJUGE OU COMPANHEIRO(A) DE SEXO DIFERENTE	<input type="checkbox"/> 12 - IRMÃO OU IRMÃ																				
<input type="checkbox"/> 03 - CÔNJUGE OU COMPANHEIRO(A) DO MESMO SEXO	<input type="checkbox"/> 13 - AVÔ OU AVÓ																				
<input type="checkbox"/> 04 - FILHO(A) DO RESPONSÁVEL E DO CÔNJUGE	<input type="checkbox"/> 14 - OUTRO PARENTE																				
<input type="checkbox"/> 05 - FILHO(A) SOMENTE DO RESPONSÁVEL	<input type="checkbox"/> 15 - AGREGADO(A)																				
<input type="checkbox"/> 06 - ENTEADO(A)	<input type="checkbox"/> 16 - CONVIVENTE																				
<input type="checkbox"/> 07 - GENRO OU NORA	<input type="checkbox"/> 17 - PENSIONISTA																				
<input type="checkbox"/> 08 - PAI, MÃE, PADRASTO OU MADRASTA	<input type="checkbox"/> 18 - EMPREGADO(A) DOMÉSTICO(A)																				
<input type="checkbox"/> 09 - SOGRO(A)	<input type="checkbox"/> 19 - PARENTE DO(A) EMPREGADO(A) DOMÉSTICO(A)																				
<input type="checkbox"/> 10 - NETO(A)	<input type="checkbox"/> 20 - INDIVIDUAL EM DOMICÍLIO COLETIVO																				

ATENÇÃO: Para o preenchimento da relação de parentesco, é necessário indicar um morador como responsável pelo domicílio, entendendo-se que é a pessoa de 12 anos ou mais de idade, cuja responsabilidade domiciliar é reconhecida pelos demais moradores.

Figura 2-4 – Questionário do Censo para mapeamento de sexo e idade.
Fonte: IBGE INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, (2020b)

Com a compreensão do censo demográfico estabelecido, a próxima etapa é explorar técnicas avançadas de *machine learning* (aprendizado de máquina) adequadas para abordar questões espaço-temporais. Estas técnicas serão pertinentes para integrar e analisar os dados previamente descritos, permitindo uma compreensão melhor das dinâmicas de suicídio no contexto brasileiro.

2.3 ESTUDO ESPAÇO-TEMPORAL

Existem diversas abordagens estatísticas e computacionais para estudos espaço-temporais. Em revisões da literatura sobre estudos espaço-temporais do suicídio, observa-se uma carência de clareza metodológica. Idealmente, uma metodologia bem-articulada deveria detalhar, por exemplo, os critérios de seleção de dados, as premissas adotadas, as etapas de pré-processamento e análise, bem como a escolha e parametrização dos modelos estatísticos ou algoritmos computacionais. No entanto, muitos trabalhos apenas mencionam superficialmente o uso de um determinado modelo ou técnica, sem fornecer o embasamento necessário ou justificar sua escolha no contexto do estudo, deixando lacunas importantes na compreensão do leitor sobre o processo adotado.

Os subitens a seguir apresentam uma discussão sobre o que é um estudo espaço-temporal no âmbito geral, não só na aplicação na área da saúde, e as abordagens estatísticas e computacionais que foram adotadas para se estudar o efeito espaço-temporal do suicídio. Importante ressaltar que, nesta seção, serão listadas apenas as técnicas estatísticas e computacionais empregadas, sem aprofundamento em sua aplicação ou justificativa. A elucidação completa sobre as técnicas estatísticas computacionais selecionadas para este trabalho, assim como a justificativa para sua escolha, serão apresentadas na seção de "Material e Métodos".

2.3.1 Visão geral

Este subitem é baseado no artigo de Atluri, Karpatne e Kumar (2018).

Em estudos espaço-temporais, além dos atributos já existentes sobre o evento que está sendo estudado, a localização e o tempo também influenciam. Dessa

forma, os eventos, as instâncias, não são independentes e identicamente distribuídos, como algumas técnicas estatísticas e de aprendizado de máquina abordam. As instâncias são relacionadas no contexto do espaço e tempo e apresentam propriedades variadas em diferentes regiões e períodos. Se esses dados forem tratados como independentes, ignorando a dependência entre eles, a capacidade de acerto e interpretação do modelo não atinge seu potencial. Se, por exemplo, para prever a energia eólica gerada por uma turbina, for considerada somente a série histórica da turbina, são perdidas todas as informações sobre turbinas em região vizinha que estão instaladas na região há mais tempo e possuem mais dados históricos para ser avaliada.

O campo de estudo que estuda esse fenômeno se chama mineração de dados espaço-temporal (MDSP – *spatio-temporal data mining* STDM).

Os dados espaço-temporais possuem duas características principais:

- 1) **Autocorrelação:** as observações feitas em locais e períodos próximos são correlacionadas entre si. Utilizar *cross validation* em dados com essas características pode ser inválido, porque os grupos de treino e teste podem estar relacionados entre si.
- 2) **Heterogeneidade:** as observações não são homogêneas ao longo do espaço-tempo. Por exemplo, observações de um satélite são diferentes no inverno e no verão. Algumas vezes, a heterogeneidade requer modelos diferentes que variam com as regiões espaço-temporais.

Em estudos espaço-temporais, há diversas formas de modelar a variável resposta e as instâncias para treinamento do modelo. Uma instância é a unidade básica do dado que o algoritmo de mineração de dados opera. Cada instância é um objeto a ser modelado.

Para ilustrar o que é uma instância, a seguir são descritos três exemplos de outras áreas que também utilizam técnicas de mineração de dados espaços temporais para estudar melhor os fenômenos dos respectivos domínios. Na área de climatologia, usualmente agrupa-se **regiões** geográficas próximas que apresentam o mesmo sinal climático ao longo do tempo. Nesse cenário, as localizações são objetos, instâncias,

e as variáveis são medidas climáticas feitas ao longo do tempo. Já em neurociência, é relevante identificar o **momento** em que uma atividade cerebral acontece no cérebro. Nesse caso, cada objeto seria um ponto no tempo e as variáveis seriam baseadas na atividade em cada localização do cérebro (variáveis espaciais). E por fim, em criminologia, no desafio de entender crimes, por exemplo, cada crime é uma instância e as variáveis são baseadas tanto no espaço, quanto no tempo.

Além da decisão de qual é o objeto a ser modelado e quais são as variáveis que devem ser consideradas nos quesitos de espaço e tempo, é necessário também escolher qual técnica melhor traduz o problema que se quer resolver. Usualmente, em problemas espaço-temporais são utilizadas técnicas de agrupamento, predição de valores, detecção de anomalia e mineração de padrões (*clustering, predictive learning, frequent pattern mining, anomaly detection, change detection, and relationship mining*). Nos subitens subsequentes, descreve-se a aplicação de cada técnica, bem como os algoritmos mais frequentemente empregados em cada técnica para cada tipo de objeto a ser modelado.

Em estudos espaço-temporais, frequentemente se utilizam cinco tipos de instâncias (unidade básica de entrada que o algoritmo utiliza na modelagem): **(1)** trajetória, **(2)** mapas espaciais, **(3)** raster, **(4)** pontos e **(5)** séries temporais. Para este estudo, focou-se especificamente em pontos e séries temporais, visto que esses tipos de instância se alinham mais adequadamente à análise do suicídio.

2.3.1.1 Pontos

A instância é representada por uma tupla que contém informações espaço-temporais de uma observação discreta e variáveis adicionais associadas com a observação. Por exemplo, pode-se utilizar como instância o ato de suicídio e agrupá-lo para encontrar instâncias anômalas.

Para este estudo, são abordadas as técnicas que respondem a duas questões. A primeira questão é: como as observações se agrupam no espaço-tempo?

Essa questão pode ser respondida com técnica de agrupamento. Ao se utilizar agrupamento, normalmente, há dois objetivos envolvidos na análise. O primeiro é descobrir grupos que possuem alta densidade de pontos espaço-temporais. Nessa abordagem, é possível descobrir surtos de doença, por exemplo. Tecnicamente, utiliza-se uma generalização de uma técnica estatística de varredura espacial proposta por Kulldorff (1997) para problemas espaço-temporais (KULLDORFF, 2001)(KULLDORFF et al., 2005). Há, também, abordagens que estendem o uso de agrupamentos espaciais, como o CLARANS (NG; HAN, 2002), para problemas espaço-temporais.

Na área da saúde, a técnica proposta por Kulldorff (1997) é usualmente utilizada. Nesta técnica, percorre-se cada ponto geográfico como centroide e avalia-se o agrupamento formado pelo centroide e os pontos geográficos que estão dentro de uma janela de formato e tamanho definidos pelo usuário. Por exemplo: o formato da janela pode ser circular com tamanho máximo 30% da população total. Se o formato da janela for cilíndrico, a análise espaço-temporal é feita, com a variável do tempo como uma dimensão a mais da janela.

O agrupamento dos pontos geográficos sendo analisado pode ter mais casos observados do que o esperado. Neste caso, verifica-se qual a probabilidade de esse evento ocorrer ao acaso. Se a probabilidade de ocorrer ao acaso for menor que 5% (ou outro padrão definido pelo usuário), entende-se que há um agrupamento de alto risco significativo perante a distribuição total do número de eventos.

Para se concluir que há mais casos observados do que esperado, faz-se o teste da razão de verossimilhança, em que a hipótese alternativa é que a probabilidade (“p”) do evento ocorrer no agrupamento (“Z”) sendo analisado é maior que a probabilidade (“q”) de ocorrer fora do agrupamento, no restante da população total. A população total é representada por “G”. Em um exemplo do suicídio, utiliza-se a distribuição de Poisson para se elaborar a função de verossimilhança.

A seguir a equação 1 apresenta a função de verossimilhança da distribuição, se “p” é diferente de “q” (hipótese alternativa), já com a média representando o “ λ ” da distribuição de Poisson (após realizar estimativa de máxima verossimilhança):

$$\begin{aligned}
L(Z, p, q) &= \frac{e^{-p\mu(Z)-q(\mu(G)-\mu(Z))} [p\mu(Z) + q(\mu(G) - \mu(Z))]^{n_G}}{n_G!} \\
&\times \prod_{x_i \in Z} \frac{p\mu(x_i)}{p\mu(Z) + q(\mu(G) - \mu(Z))} \prod_{x_i \notin Z} \frac{q\mu(x_i)}{p\mu(Z) + q(\mu(G) - \mu(Z))} \\
&= \frac{e^{-p\mu(Z)-q(\mu(G)-\mu(Z))}}{n_G!} p^{n_Z} q^{(n_G - n_Z)} \prod_{x_i} \mu(x_i)
\end{aligned} \tag{1}$$

onde:

- p = probabilidade de o evento ocorrer dentro do agrupamento;
- Z = agrupamento que está sendo analisado;
- q = probabilidade de o evento ocorrer fora do agrupamento;
- G = população total;
- μ = tamanho do espaço a ser determinado, por exemplo, " $\mu(Z)$ " representa o tamanho da população no agrupamento " Z ";
- η = número de eventos dentro do espaço a ser determinado, exemplo " ηG " representa o número de eventos (suicídio) que ocorreram no espaço " G ";
- x_i = cada ponto geográfico existente no mapa.

Se a probabilidade de " p " é igual à probabilidade de " q " (hipótese nula), então a função de verossimilhança é apresentada na equação 2.

$$L_0 = \sup_p \frac{e^{-p\mu(G)} p^{n_G}}{n_G!} \prod_{x_i} \mu(x_i) = \frac{e^{-n_G}}{n_G!} \left(\frac{n_G}{\mu(G)} \right)^{n_G} \prod_{x_i} \mu(x_i) \tag{2}$$

A razão de verossimilhança é ilustrada na equação 3.

$$\lambda = \frac{\sup_{Z \in Z} L(Z)}{\frac{e^{-n_G} \left(\frac{n_G}{\mu(G)}\right)^{n_G}}{n_G!} \prod_{x_i} \mu(x_i)} \quad (3)$$

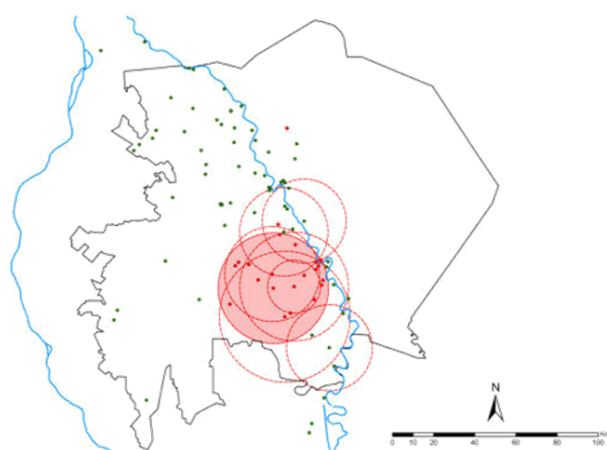
$$= \sup_{Z \in Z} \frac{\left(\frac{n_Z}{\mu(Z)}\right)^{n_Z} \left(\frac{n_G - n_Z}{\mu(G) - \mu(Z)}\right)^{n_G - n_Z}}{\left(\frac{n_G}{\mu(G)}\right)^{n_G}} I\left(\frac{n_Z}{\mu(Z)} > \frac{(n_G - n_Z)}{(\mu(G) - \mu(Z))}\right)$$

onde:

- I = função identidade.

A função de identidade é igual a 1 se houver ao menos um agrupamento em que há mais eventos proporcionalmente ao total da população sem considerar o agrupamento. Em casos em que se deseja obter os agrupamentos de baixo risco, a mesma lógica matemática é realizada, mas analisando agrupamentos em que a proporção do evento é menor do que no restante da população.

Para se garantir que a proporção encontrada é significativa, realiza-se o teste de hipótese de Monte Carlo. No teste de Monte Carlo, além da distribuição de eventos que ocorre (eventos observados), conjuntos de dados aleatórios são gerados considerando a distribuição de Poisson original do dado e analisa-se a classificação da maior razão de verossimilhança para a menor. Se o agrupamento do conjunto real de dados estiver dentro das 5% maiores razões de verossimilhança, ele é considerado significativo porque apresenta um p-valor menor que 5%. O número de simulações normalmente termina-se com 999 para facilitar o cálculo matemático (KULLDORFF, 1997) pois o p-valor é calculado a partir da classificação C do conjunto observado, em que $p\text{-valor} = C / (1 + \#\text{simulações})$. A Figura 2-5 resume o algoritmo de agrupamento espaço ou espaço-temporal proposto por Kulldorff.



Fonte: Research Gate

1. Cada ponto geográfico centróide e avalia-se o agrupamento formado.
2. Se o agrupamento tiver mais casos observados que esperado, verifica-se a probabilidade de esse evento ocorrer ao acaso.
3. Conjuntos de dados aleatórios são gerados considerando a distribuição de Poisson original do dado e analisa-se a classificação da maior razão de verossimilhança para a menor. (Simulação Monte Carlo)

Figura 2-5 – Resumo ilustrativo do algoritmo proposto por Kulldorff.

A segunda questão é: é possível estimar uma variável resposta em um ponto no domínio espaço-temporal?

O segundo objetivo da utilização de técnicas espaço-temporais em pontos é encontrar pontos que apresentam características não espaço-temporais similares, como encontrar pontos com taxas de suicídio semelhantes. Estatisticamente, uma abordagem possível é fazer uma extensão do algoritmo DBSCAN, um algoritmo de agrupamento (*clustering*) baseado em densidade, para problemas espaço-temporais, como o ST-DBSCAN (BIRANT; KUT, 2007). Na proposta do ST-DBSCAN, as distâncias espaciais, temporais e não espaço-temporais são calculadas separadamente, e há limites numéricos diferentes para cada distância. Ainda é um desafio lidar com a heterogeneidade do espaço e tempo, e com as densidades de grupos com bastante variância.

Outra forma de se responder à questão seria com técnica de classificação. Um exemplo em epidemiologia seria prever a taxa de suicídio de um local baseada nas taxas e características de localizações e tempos vizinhos. Modelos de causalidade espaço-temporal de Granger consideram a relação espaço-temporal dos dados, utilizando dados espaço-temporais na regressão (LOZANO et al., 2009)(LUO et al., 2013). Modelos tradicionais espaciais algumas vezes resolvem o problema, mas não apresentam a correlação temporal. São eles: modelos espaciais autorregressivos (*spatial auto-regressive SAR*) (KELEJIAN; PRUCHA, 1999), regressão ponderada geograficamente (*geographically weighted regression GWR*) (BRUNSDON;

FOTHERINGHAM; CHARLTON, 1998) e *Kriging* ou krigagem (OLIVER; WEBSTER, 1990).

Os algoritmos anteriormente citados não são detalhadamente discutidos, uma vez que não são empregados nesta pesquisa. Contudo, a vastidão de opções algorítmicas disponíveis para a condução de estudos torna evidente a relevância de comparar todos os algoritmos potenciais usando as mesmas bases de dados e métricas. Essa abordagem comparativa é relevante para embasar decisões informadas sobre a escolha do algoritmo mais adequado.

2.3.1.2 Séries temporais

Em desafios em que a unidade básica que o algoritmo de aprendizagem estatística opera é a série temporal, é comum querer responder às questões: é possível identificar grupos de séries temporais que apresentam atividade temporal similar e são localizados próximos? Há algum padrão temporal que comumente se repete em um número de séries temporais? É possível identificar momentos em que a série apresenta uma mudança em seu perfil?

A primeira questão pode ser respondida com técnica de agrupamento. Em agrupamento, o objetivo principal de agrupar séries temporais é encontrar grupos de localidade com atividades temporais similares. Uma abordagem é agrupar a série com técnicas tradicionais e, após essa etapa, trabalhar com o fator espacial, como aumentar o número de grupos resultantes para que as unidades sejam vizinhas espacialmente (SMITH et al., 2009). Por exemplo, inicialmente, é possível empregar técnicas de agrupamento para organizar os casos de suicídio conforme o perfil dos indivíduos, como, por exemplo, homens jovens. Posteriormente, esses conjuntos podem ser subdivididos com base na localização residencial dos indivíduos, originando agrupamentos adicionais mais específicos.

Em detecção de padrões, pode-se encontrar segmentos das séries temporais que se repetem em localizações múltiplas. Por exemplo, na agricultura, é possível identificar o ciclo de colheita quando muitas fazendas com o mesmo plantio realizam a colheita. Há abordagens que não consideram a relação espacial, como as propostas

por Chiu, Keogh e Lonardi (2003), Meng et al. (2008), Minnen et al. (2007), Yeh et al. (2017), Zhu et al. (2017) e as que já consideram a dinâmica espaço-temporal, como os propostos por Takahashi et al. (2008), Zhao et al. (2015), Zhou e Matteson (2015).

Na previsão de mudança de padrão, o objetivo é identificar em qual momento o sistema passa por um desvio significativo em relação ao seu passado. Atluri, Karpatne e Kumar (2018) apresentam uma revisão bibliográfica do assunto que considera mais de 20 artigos.

2.3.2 Estudo espaço-temporal do suicídio

Foi realizada uma pesquisa baseada na metodologia de revisão de escopo proposta por Tricco et al. (2018) com o objetivo de se identificar quais são os métodos e técnicas aplicadas em estudos espaço-temporais do suicídio.

Na revisão, foram levantados os artigos, em inglês, português ou espanhol, das bases eletrônicas indexadas Pubmed (2022), IEEE Explore (2022), ACM (2022) e Web of Science (2022), que apresentassem estudo espaço-temporal do ato do suicídio, publicados entre 2017 e 2020. Foram considerados artigos que descrevessem o método no resumo e que possuíssem ilustrações de mapas ou tabelas no texto. Foram encontrados 99 artigos na plataforma do Pubmed (2022), dos quais 16 foram selecionados por atingirem o critério de sucesso da revisão. Não foram encontrados artigos para metodologias de estudo de suicídio nas outras bases eletrônicas indexadas que fossem diferentes dos artigos já encontrados.

A chave de busca foi: (geospatial OR spat* OR spat* epidemiology OR spat* temporal OR spatial-temporal OR spatio-temporal OR spat* cluster* OR spat* analys* OR georeference OR geodata OR geoanalytics OR temporal OR space-time OR time series) AND (suicide*). Também foram adicionados dois artigos de 2012 por indicação do grupo de pesquisa.

Apesar de não mencionarem explicitamente as etapas adotadas no estudo, encontra-se presente a seguinte metodologia nos artigos selecionados:

1. Definição dos dados: definição dos dados que foram utilizados no estudo, com definição das fontes e granularidades geográficas e temporais.
2. Método do cálculo da taxa de suicídio: há diversas formas de se calcular a taxa.
3. Análise descritiva: análises iniciais cujo objetivo é verificar a qualidade dos dados e trazer as primeiras conclusões.
4. Análise espacial: correlação entre o espaço geográfico e o suicídio.
5. Análise temporal: correlação entre o tempo e o suicídio.
6. Análise multivariada: além do espaço e tempo, alguns artigos adicionam algumas variáveis para melhor entendimento do tema.

Os autores usualmente não mencionam as limitações das metodologias estatísticas e computacionais adotadas em cada etapa, então recomenda-se como trabalho futuro adotar a metanálise proposta por Egger, Smith e Altman (2008) para se entender a limitação de cada abordagem realizada e qual abordagem estatística foi mais eficaz para o estudo em questão. Na metanálise mencionada, é possível comparar a métrica de sucesso entre artigos que utilizaram bases e algoritmos estatísticos e computacionais diferentes entre si.

Apesar de não haver uma comparação quantitativa dos métodos realizados, os subitens a seguir apresentam um paralelo entre as principais metodologias adotadas na literatura para cada etapa de estudo e na seção “Material e métodos” é descrito o método utilizado para a pesquisa e justificativa.

2.3.2.1 Definição dos dados

Além dos dados geográficos e temporais, normalmente os artigos também procuram correlacionar o suicídio com outras variáveis que buscam representar os fatores de risco comentados no “INTRODUÇÃO E JUSTIFICATIVA” com a granularidade da unidade espaço-temporal a ser estudada.

Relembrando, os fatores de risco são: (1) riscos relacionados às comunidades e aos relacionamentos, (2) riscos sociodemográficos, (3) riscos relacionados ao nível individual e (4) riscos ambientais.

Sobre os riscos relacionados às comunidades e os sociodemográficos, Qi, Tong e Hu (2009), Fontanella et al. (2018), Yu et al. (2020), Santos, Barbosa e Severo (2020), Yamaoka et al. (2020), Gowda et al. (2019), Alarcão et al. (2019), Sy et al. (2019), Haghparast-Bidgoli et al. (2018), Rossen et al. (2018) e Knipe et al. (2017) utilizam variáveis do Censo para entender esse tópico. As variáveis utilizadas nos estudos englobam: tamanho da população, distribuição da população por gênero masculino/feminino e idade, etnia, nível de renda, nível educacional, grau de desemprego, estado civil, percentual que mudou de casa no último ano, percentual de população que vive em zonas rurais, tamanho da família, grau de urbanização e taxa de migração.

Ainda no entendimento das variáveis socioeconômicas e demográficas, Qi, Tong e Hu (2009), Bando et al. (2012), Tran e Morrison (2020) e Santos, Barbosa e Severo (2020) utilizam índices definidos a partir do Censo, como índices relacionados à desigualdade socioeconômica (Gini Index), índice de desenvolvimento humano (IDH), produto interno bruto (PIB) per capita e renda média.

Outra abordagem para se entender socioeconomicamente a região, seria a proposta por Gowda et al. (2019) e Rossen et al. (2018) que analisam a quantidade de policiais existentes e quantidade de crimes em cada unidade de estudo para caracterizar socioeconomicamente a região.

Sobre os riscos relacionados ao nível individual, os estudos consideram os dados anonimizados da vítima de suicídio. Normalmente, junto com a base de dados do ato, também há informações de gênero masculino/feminino, idade, estado civil, local de residência, método de suicídio e data. (QI, TONG e HU, 2009) (BANDO et al., 2012) (SAMAN et al., 2012) (FONTANELLA et al., 2018) (TRAN e MORRISON, 2020) (YU et al., 2020) (SANTOS, BARBOSA e SEVERO, 2020) (PALMA, DOS SANTOS e IGNOTTI, 2020) (YAMAOKA et al., 2020) (DOS SANTOS et al., 2018).

Riscos no nível individual consideram também transtornos mentais. Por isso, Fontanella et al. (2018) buscam caracterizar a unidade de estudo pelo número de centros de Saúde Mental no local, número de leitos psiquiátricos, taxa de atenção médica primária e taxa de psiquiatras presentes na unidade. Yamaoka et al. (2020) consideram em seu estudo o número de hospitais psiquiátricos presentes na região. Rossen et al. (2018) consideram dados de consumo de drogas e dados de recursos médicos.

Os fatores de risco ambientais também foram estudados por Qi, Tong e Hu (2009) e Yu et al. (2020). Os primeiros autores utilizam informação de precipitação, temperatura máxima e mínima. Já os segundos autores estudam a poluição do ar e indicadores climáticos, como a concentração média anual de partícula fina, medições do nível de dióxido e de nitrogênio e temperatura média diária.

Após esta etapa de seleção dos dados, os artigos explicam como foi feito o cálculo da taxa de suicídio.

2.3.2.2 Método do cálculo da taxa de suicídio

O suicídio é considerado um evento raro. Dependendo da granularidade espaço-temporal escolhida no estudo, o número de eventos ocorridos pode não ser estatisticamente significativo para ser considerado e há a possibilidade de gerar uma taxa instável com alta variância. Por exemplo, um município que apresentou em um ano uma morte por suicídio e, no ano seguinte, apresentou duas mortes, dobrou as mortes por suicídio de um ano para o outro. (ROSSEN et al., 2018)

Para mitigar essa instabilidade, há artigos que realizam a suavização espacial feita por modelos bayesianos na taxa de suicídio (FONTANELLA et al., 2018) (PALMA, DOS SANTOS e IGNOTTI, 2020) (YAMAOKA et al., 2020) (ALARCÃO et al., 2019) (ROSSEN et al., 2018). O método Bayesiano é indicado quando há áreas com populações reduzidas. Ele estima uma taxa, considerando o valor observado naquela granularidade geográfica e em seus vizinhos. A suavização espacial pode ser feita utilizando o pacote *diseasemapping* do R (FONTANELLA et al., 2018), o pacote INLA

do R (ROSSEN et al., 2018), o *software* TerraView (PALMA, DOS SANTOS e IGNOTTI, 2020) e o *software* SAS (YAMAOKA et al., 2020).

Rossen et al. (2018) propõem alternativas a esse tema também no cálculo da taxa de suicídio: desconsiderar regiões que apresentam menos que 20 mortes por suicídio e/ ou agrupar os anos de estudo para se evitar a ocorrência de evento raro no estudo.

Como o objetivo dos estudos espaço-temporais é entender ainda mais o suicídio, é interessante retirar esses vieses de base já conhecidos, para ser possível entender os riscos dos outros fatores que estão sendo estudados. É de conhecimento da comunidade científica que a taxa do suicídio varia pelos estágios da vida (faixa etária) e é diferente para homens e mulheres (WHO, 2016). Por isso, um dos métodos do cálculo da taxa de suicídio é ajustar a taxa calculada por idade e gênero masculino/feminino (cálculo ponderado) (QI, TONG e HU, 2009) (SANTOS, BARBOSA e SEVERO, 2020) (BANDO et al., 2012), (SAMAN et al., 2012) (TRAN E MORRISON, 2020) (YU et al., 2020) (SY et al., 2019) (SANTOS et al., 2018).

Por fim, há artigos que fazem o cálculo simples da taxa de suicídio, que é o número de mortes/total da população da região (GOWDA et al., 2019) (HAGHPARAST-BIDGOLI et al., 2018) (LOUREIRO et al., 2018) (KNIPE et al., 2017).

Uma vez obtidos os dados e definida a forma de cálculo, segue-se para a análise descritiva dos dados.

2.3.2.3 Análise Descritiva

Na etapa de análise descritiva, os estudos geralmente possuem dois objetivos: analisar a qualidade dos dados e realizar uma análise prévia da correlação espaço-temporal e outras variáveis abordadas.

Objetivando verificar a qualidade dos dados, Qi, Tong e Hu (2009), Bando et al. (2012), Tran e Morrison (2020), Yu et al. (2020), Yamaoka et al. (2020) e Palma,

Santos e Ignotti (2020) realizam a análise univariada das variáveis do estudo (contagem, média, máximo e mínimo).

Para se entender preliminarmente a correlação temporal, Yamaoka et al. (2020) fazem histogramas da taxa de suicídio e Yu et al. (2020), Yamaoka et al. (2020) e Rossen et al. (2018) analisam as tendências históricas.

No entendimento inicial da correlação espacial, Bando et al. (2012), Gowda et al. (2019) e Rossen et al. (2018) fazem mapas geográficos com a distribuição da taxa de suicídio, e Alarcão et al. (2019) fazem o teste do coeficiente bivariado de Moran I, para correlacionar o suicídio, o espaço e as variáveis independentes do estudo.

Para entender a correlação primária do suicídio com outras variáveis, além das espaciais e temporais, Qi, Tong e Hu (2009) e Saman et al. (2012) realizam a correlação de Pearson e Teste T entre a taxa de suicídio e as covariáveis do estudo. Sy et al. (2019) e Haghparast-Bidgoli et al. (2018) apresentam tabelas e *boxplots* comparativos que mostram como o suicídio (média e desvio-padrão) se distribui de acordo com cada variável independente, incluindo o espaço geográfico e o tempo.

Portanto, dos estudos analisados, verifica-se que há diversas formas de se fazer o levantamento inicial das hipóteses.

Seguindo o que os artigos fazem, após a análise prévia das hipóteses presente na análise descritiva, nas próximas seções são aprofundadas as hipóteses com ferramental estatístico e computacional mais complexo.

2.3.2.4 Análise Espacial

Na análise espacial, o objetivo é entender se há correlação espacial com o suicídio. Deseja-se entender, primeiramente, se o suicídio se associa no espaço geográfico ou se ele se apresenta aleatoriamente no espaço. Isso é chamado de índice de correlação espacial. Para a realização dessa etapa, Qi, Tong e Hu (2009) utilizam técnicas de análise de semivariograma com os *softwares* SPSS e +SpatialStats. O teste de Moran Global I ou Global também apresenta o mesmo

resultado. Tran e Morrison (2020) utilizam o *software* R e Santos, Barbosa e Severo (2020), Grigoroglou et al. (2019) e Santos et al. (2018) utilizam o *software* Terraview.

Além disso, deseja-se também entender se há algum agrupamento de unidades geográficas em que o fator de risco é maior, e qual é esse agrupamento. Isso é feito através do agrupamento espacial. Foram mapeadas quatro técnicas utilizadas: modelo de Kulldorff (KULLDORFF, 1997), baseado em Poisson, utilizando o *software* SaTScan (BANDO et al., 2012) (SAMAN et al., 2012) (FONTANELLA et al., 2018) (PALMA, SANTOS e IGNOTTI, 2020) (YAMAOKA et al., 2020) (SY et al., 2019) (LOUREIRO et al., 2018); teste de Moran Local, utilizando o *software* GeoDa e QGIS (ALARCÃO et al., 2019); modelo de varredura espacial flexível de Tango-Takahashi, utilizando o *software* FlexScan (YAMAOKA et al., 2020); e teste de Tango, utilizando o *software* Disease Mapping System (YAMAOKA et al., 2020).

Há artigos que apresentam evidências suficientes de associação e agrupamento espacial somente com a observação do mapa com a distribuição geográfica do suicídio, como é mostrado em Qi, Tong e Hu (2009), Saman et al. (2012), Yu et al. (2020), Grigoroglou et al. (2019), Rossen et al. (2018), Knipe et al. (2017).

Para a visualização dos mapas, foram utilizadas as ferramentas Mapinfo (QI, TONG e HU, 2009), ArcGIS (BANDO et al., 2012) (SAMAN et al., 2012) (FONTANELLA et al., 2018) (KNIPE et al., 2017), R pacote QGIS (TRAN e MORRISON, 2020) (ALARCÃO et al., 2019), R pacote ggmap (SY et al., 2019), Terraview (SANTOS, BARBOSA e SEVERO, 2020), GeoDa (ALARCÃO et al., 2019) e ArcMap (LOUREIRO et al., 2018).

Após o entendimento espacial, no próximo item é descrito como na literatura se aborda a análise temporal

2.3.2.5 Análise Temporal

Na análise temporal, um dos objetivos é entender quais unidades geográficas de estudo apresentam tendência de variação temporal da taxa de suicídio. Para a

análise temporal, os autores dos artigos selecionados adotam diferentes técnicas, mencionadas na sequência.

Em uma abordagem mais simples, é realizada a observação do histórico dos mapas e linhas de tendência nos gráficos. (QI, TONG e HU, 2009) (PALMA, SANTOS e IGNOTTI, 2020) (ALARCÃO et al., 2019) (ROSSEN et al., 2018) (KNIPE et al., 2017).

Grigoroglou et al. (2019) e Rossen et al. (2018) realizam a correlação de Spearman entre a taxa de suicídio e o ano. Para se entender se houve mudança da tendência histórica comparado entre as cidades, Rossen et al. (2018) comparam a ordenação das cidades com maiores taxas em um ano com relação à ordenação 10 anos depois (ex.: comparar classificação de 2005 com 2015).

É interessante considerar a correlação de Spearman para se analisar o crescimento da taxa de suicídio porque esta abordagem estatística considera todo o histórico em seu cálculo. Dessa forma, crescimento contínuo do objetivo em questão é diferente do crescimento do objeto apenas nos últimos anos, por exemplo. Saber se a mortalidade por suicídio vem aumentando continuamente ou se é um acontecimento que acelerou apenas nos últimos anos é importante para o especialista de Saúde Pública.

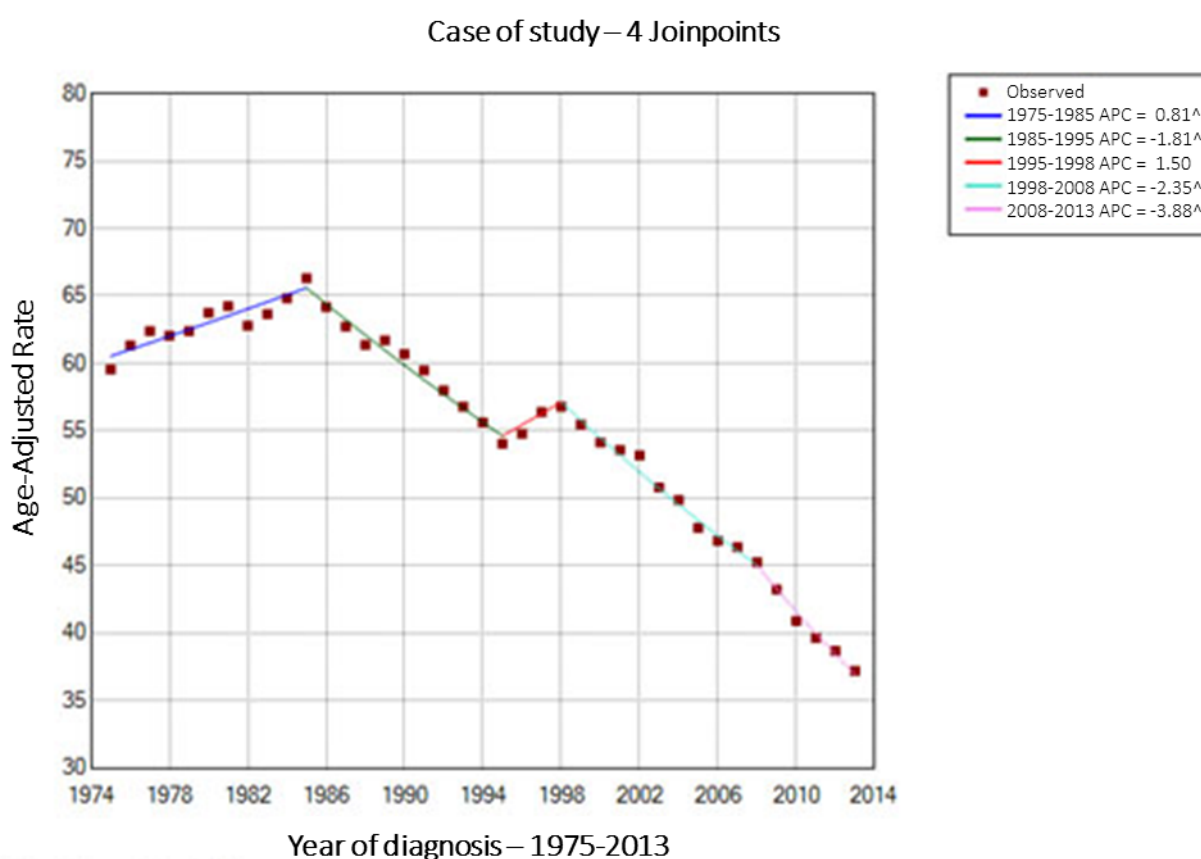
Gowda et al. (2019) e Santos et al. (2018) realizam uma regressão linear em que o ano é a variável independente, e a variável dependente é a taxa de suicídio. Gowda et al. (2019) utilizam o *software* SPSS. Santos et al. (2018) utilizam GraphPad Prism e Microsoft Excel. O ponto de atenção desta técnica é que considera somente crescimentos lineares e contínuos.

Bando et al. (2012), Santos, Barbosa e Severo (2020) e Loureiro et al. (2018) utilizam a regressão de Joinpoint (modelo *log-linear*, que utiliza regressão de Poisson), que identifica pontos de inflexão nas tendências da taxa de suicídio, oferecido pelo *software* Joinpoint Regression Program (NATIONAL CANCER INSTITUTE, SURVEILLANCE RESEARCH PROGRAM, STATISTICAL METHODOLOGY AND APPLICATIONS BRANCH), (2022). Como essa regressão identifica pontos de

inflexão, é possível observar tendências de crescimento em faixas diferentes de tempo.

Esse *software* foi desenvolvido pelo programa de pesquisa de vigilância do câncer, um programa pertencente ao Instituto Nacional do Câncer dos Estados Unidos (*National Cancer Institute* – NCI). Esse instituto é o principal polo de pesquisa e treinamento sobre câncer do governo estado-unidense (NCI, 2018). Sua metodologia temporal de análise é difundida, tanto que se vê, na pesquisa bibliográfica, grande parte dos artigos utilizando suas técnicas.

Na regressão de Joinpoint, o *software* utiliza o conceito de “*segmented regression*”, “*piecewise regression*” ou “*broken-stick regression*”. Nessa metodologia, a variável independente é quebrada em intervalos e, para cada intervalo, tenta-se calcular a melhor regressão para obtenção do menor erro (KIM et al., 2000). Na Figura 2-6 é ilustrado um exemplo da saída de uma análise de Joinpoint.



[^]The Annual Percent Change (APC) is significantly different from zero at $\alpha = 0.05$

Figura 2-6 – Exemplo da saída do *software* Joinpoint.
Fonte: NATIONAL CANCER INSTITUTE (2022)

Para o caso do estudo do suicídio, os intervalos são os anos de análise e tenta-se entender se há uma relação linear entre o logaritmo da taxa de suicídio de cada município, ajustada por idade e gênero masculino/feminino, ao longo dos anos. Se há uma relação linear entre os logaritmos da taxa ajustada, significa que a taxa muda em um percentual constante por ano (KIM et al., 2000).

Para se descobrir a melhor localização dos intervalos, uma das metodologias é utilizar *grid-search* que combina todos os intervalos possíveis para a quantidade de cortes/quebras que está sendo analisada. Para seleção de qual o número de quebras/cortes que é adotado, utiliza-se a permutação com amostragem Monte Carlo.

O teste de permutação compara dois modelos que apresentam um número de corte diferente entre si. Um dos modelos, que representa a hipótese nula, é o mais simples, com menos corte que o segundo modelo, a hipótese alternativa. Analisa-se se a razão entre a soma do erro quadrático médio do modelo da hipótese nula e a soma do erro quadrático médio do modelo da hipótese alternativa é diferente de 1. Se a razão for diferente de 1, há indícios para se recusar a hipótese nula. Para se entender se a razão é diferente de 1 com significância estatística, ou seja, não é um acontecimento do acaso, os erros/resíduos do conjunto de dados modelado para a hipótese nula são permutados e verifica-se a razão para as inúmeras permutações feitas. Se a razão da amostra observada for raro de acontecer, isto é, nas classificações das permutações, observa-se que ela acontece em menos de 5% das vezes (conceito do p-valor), rejeita-se a hipótese nula e considera-se que o modelo deve ter o número de quebras/cortes definido pela hipótese alternativa (KIM et al., 2000).

O programa Joinpoint aponta quais são as séries temporais que possuem a taxa de mudança anual (*Annual Percent Change - APC*) significativa, ou seja, considerando um alpha de 5%, há têm uma nova tendência temporal que é diferente da hipótese nula. Porém, há casos em que ele aponta que há uma tendência histórica, mas, quando se analisa visualmente a série, não é possível ver essa tendência. Isso pode acontecer porque há poucos pontos, o que pode reduzir o poder do teste estatístico para se calcular a significância. Esse efeito deve ser considerado caso se

utilize a regressão de Joinpoint nas análises. A Figura 2-7 resume o passo a passo desse método.

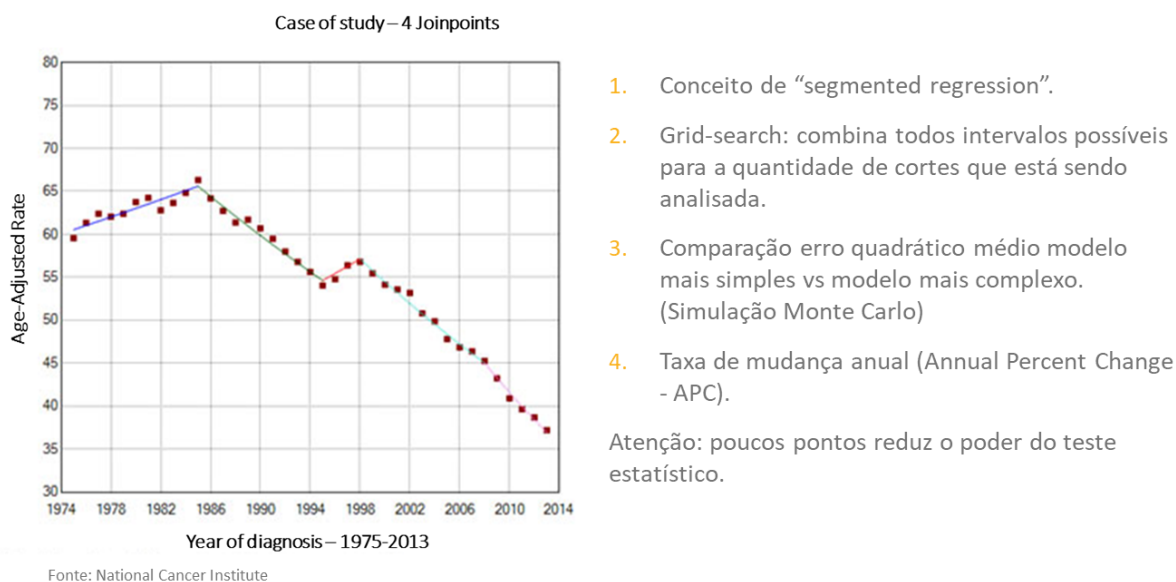


Figura 2-7 – Passo a passo resumido da regressão de Joinpoint.

Além da técnica de Joinpoint, Saman et al. (2012), Fontanella et al. (2018), Yamaoka et al. (2020) e Sy et al. (2019) utilizam técnicas de agrupamento temporal ou espaço-temporal elaboradas por Kuldorff (2005) e disponíveis no *software* SaTScan, para descobrir unidades temporais que apresentam alto fator de risco para o suicídio. Se os últimos anos do estudo estiverem em um desses agrupamentos, significa que há a probabilidade de o próximo ano também ser um ano de alto risco.

Outro objetivo da análise temporal é entender se há sazonalidade do suicídio entre os meses e dias do ano. Yu et al. (2020) fazem essa análise por meio da estimativa da sazonalidade e uso de metarregressão para agrupar valores. Eles utilizam os pacotes no R *stl* e *mixmeta*.

Finalizando a etapa temporal, os artigos usualmente investigam a correlação do suicídio com outros fatores de risco através da análise multivariada.

2.3.2.6 Análise Multivariada

Na análise multivariada, busca-se entender a correlação de todos os fatores de risco apontados no subitem “2.3.2.1 – Definição dos dados” com o suicídio. Há artigos que utilizam técnicas mais simples, como interpretação dos mapas e comparação entre as taxas médias de suicídio obtidas no agrupamento. (PALMA, SANTOS e IGNOTTI, 2020) (SY et al., 2019)

É possível também realizar testes estatísticos, como testes G, qui-quadrado de Pearson e Teste T bicaudal, para interpretar os agrupamentos. Saman et al. (2012) utilizaram o *software* WinPepi. Fontanella et al. (2018) utilizaram o SAS e Santos et al. (2018) utilizaram GraphPad Prism e Microsoft Excel. Yu et al. (2020) realizam teste de Wald para entender se há diferença entre os subgrupos da análise.

Santos, Barbosa e Severo (2020) realizam a análise bivariada utilizando indicadores locais de associação espacial (*Local Indicators of Spatial Association – LISA*), utilizando o *software* GeoDa. Gowda et al. (2019) realizam a correlação de Spearman para cada variável independente e a taxa de suicídio.

Técnicas de regressão são comumente utilizadas na análise multivariada. É possível fazer uma regressão logística para interpretar os agrupamentos com alto risco. Bando et al. (2012) utilizam o *software* SPSS para isso. Fontanella et al. (2018) também utilizam a regressão logística, mas realizaram análise de componente principal antes, para reduzir a dimensionalidade de suas variáveis independentes.

Haghparast-Bidgoli et al. (2018) utilizam modelo de decomposição (*decomposition model*) com as variáveis independentes.

Knipe et al. (2017) fazem a correlação de Spearman das variáveis independentes e a taxa de suicídio e, após essa correlação, fazem uma regressão binomial negativa, para entender melhor o efeito das variáveis independentes que tiveram correlação com o teste feito anteriormente.

Qi, Tong e Hu (2009) realizam uma equação de estimativa generalizada (*generalized estimating equation – GEE*) para distribuição de Poisson (SPSS +SpatialStats softwares).

Para finalizar as regressões, Tran e Morrison (2020) e Yamaoka et al. (2020) comparam regressões para avaliar qual o modelo que interpreta melhor os resultados. Tran e Morrison (2020) comparam regressão com mínimos quadrados (base), regressão multinível e regressão ponderada geograficamente (*Geographically Weighted Regression* - GWR), utilizando o *software* R e os pacotes stats, lme4 e GWModel. Yamaoka et al. (2020) comparam regressão linear e regressão autorregressiva condicional (*conditional autoregressive* - CAR).

Alarcão et al. (2019) apresentam uma nova abordagem através de Redes Bayesianas e análise de dois especialistas no assunto, para entender relações causais, utilizando o pacote bnpa do R.

Assim como no item “2.3.2.3 Análise Descritiva”, na análise multivariada, observa-se uma variedade de ferramentas e de metodologias estatísticas e computacionais disponíveis.

Após esse levantamento de metodologias, no próximo capítulo são recomendadas as abordagens estatísticas e computacionais para o presente estudo.

3 MATERIAL E MÉTODOS

Baseado na revisão bibliográfica de estudos espaço-temporais do suicídio, esta pesquisa divide a metodologia em dois blocos principais, com quatro etapas cada bloco, são eles:

1. Dados:
 - a. Definição das fontes de dados, captura e transformação dos dados
 - b. Validação da qualidade dos dados puros/crus.
 - c. Definição do método de cálculo da taxa de suicídio.
 - d. Definição dos métodos de validação dos dados e de tratamento das variáveis (preparação dos dados).
2. Análise: definição da abordagem estatística e computacional, ferramenta, critério de sucesso, aplicação e avaliação das análises:
 - a. Descritiva.
 - b. Espacial.
 - c. Temporal.
 - d. Multivariada.

No item 1, as bases do estudo espaço temporal do suicídio e seus metadados estão no repositório Dada, 2023. Esse repositório encontra-se dentro do domínio da Universidade de Havard do projeto *Dataverse* (DATAVERSE, 2023). A Universidade de Havard disponibiliza um repositório aberto a todos os pesquisadores do mundo para depósito de bases de dados. Nessa plataforma, é gerada uma citação formal das bases de dados inseridas na mesma, com a identificação oficial do objetivo (*Digital Object Identifier – DOI*). Dessa forma, os dados se tornam mais facilmente encontráveis, iniciando o movimento de gestão dos dados inspirado no princípio denominado *Findable, Accessible, Interoperable and Reusable (FAIR)*. Em 2016, um grupo de representantes da academia, indústria, agências fomentarias e editoras acadêmicas determinaram uma série de princípios de gestão de dados para impulsionar o reuso de dados acadêmicos. Esse grupo determinou que os dados devem ser encontráveis, acessíveis, interoperáveis e reutilizáveis (WILKINSON et al., 2016).

Sobre o item 2, cada análise da metodologia tem o objetivo de investigar um ou mais pilares de questionamentos a que este estudo é direcionado. Os pilares estão descritos no CAPÍTULO 1, subitem “1.1 – OBJETIVOS”.

Na análise descritiva, espera-se investigar, por meio de gráficos, tabelas e mapas, uma visão preliminar das quatro áreas de questionamento que direcionam o estudo.

Na análise que enfatiza a questão espacial, espera-se aprofundar no pilar: “Quais agrupamentos espaciais podem estar relacionados com os dados sobre suicídio? Esses agrupamentos mudam ao longo do tempo? Há algum agrupamento de maior relevância?”.

Na análise que enfatiza a questão temporal, espera-se aprofundar-se nos pilares: “Quais municípios e quais grupos de gênero masculino/feminino e idade contribuem para o aumento da taxa de suicídio no Brasil ao longo do tempo? Historicamente, quais são as cidades com maiores crescimentos da taxa e as com maiores quedas?”.

Na análise multivariada, espera-se aprofundar-se nos pilares: “O percentual de urbanização contribui para explicar as variações das taxas de suicídio?”.

A

Tabela 3.1 resume as oito etapas, o que foi levantado de técnicas na revisão bibliográfica de cada uma delas e o que é proposto para a análise final deste projeto com as ferramentas utilizadas. Os itens a seguir descrevem cada etapa.

	Abordagem Técnica	Revisão bibliográfica estudo espaço-temporal		Proposta estudo	
		Suicídio	Geral		
Dados	Fontes de dados	<ul style="list-style-type: none"> - Variáveis sociodemográficas do censo - Índices - Recursos médicos do local. - Dados ambientais e meteorológicos - Dados de nível individual da vítima 	Não se aplica	SIM/ Tabnet (Mortalidade) e IBGE	
	Variáveis			- Taxa ajustada por idade e sexo	- Informações da vítima: data do óbito, município residência, causa do óbito, sexo, idade e raça
	Método de cálculo da taxa de suicídio			- Suavização espacial (modelo bayesiano empírico local) Ferramentas: diseasemapping no R, TerraView e SAS	- Tamanho da população por idade e sexo
	Validação dos dados e tratamento variáveis			- Tratamento de anomalia: média, máximo e mínimo, quartis - Tratamento de campos vazios	- Definição de área urbana ou rural feita pelo IBGE
				- Taxa ajustada por idade e sexo	
				Ferramenta: Python	
				- Tratamento de anomalia: média, máximo e mínimo, quartis	
				- Agrupamento das taxas de suicídio para se entender casos anômalos (<i>outliers</i>)	
				- Validação de duplicação de bases	
				- Tratamento de campos vazios e faltantes	
				Ferramenta: Python	
Análise	Descritiva	<ul style="list-style-type: none"> - Histogramas das variáveis de interesse - Testes de correlação da taxa de suicídio com as variáveis a serem estudadas - Visualização nos mapas da distribuição das variáveis - Análise visual de tendências históricas no gráfico Ferramentas: Mapinfo, ArcGIS, QGIS do R e Terraview	Não se aplica	<ul style="list-style-type: none"> - Visualização nos mapas da taxa de suicídio nos municípios - Análise visual de tendências históricas no gráfico da taxa de suicídio, com abertura de sexo e idade Ferramenta: Python	
	Espacial	<ul style="list-style-type: none"> - Índice de correlação espacial: análise de semivariograma, teste de Moran I ou Global - Agrupamento espacial: Kulldorff, varredura espacial flexível de Tango-Takahashi, teste de Tango Ferramentas: SPSS, SpatialStats, SaTScan, FlexScan, Disease Mapping System.	-Agrupamento espacial: Kulldorff, CLARANS, DBSCAN	- Agrupamento: Kulldorff Ferramenta: SaTScan e Python	
	Temporal	<ul style="list-style-type: none"> - Entender sazonalidade e tendência: decomposição da série utilizando Loess - Detectar mudança padrão: Regressão de Joinpoint - Agrupamento temporal ou espaço-temporal: Kulldorff Ferramentas: Joinpoint Regression Program, SaTScan, pacote stl e mixmeta no R	<ul style="list-style-type: none"> - Agrupar séries: k-means, agrupamento hierárquico, SNN clustering e normalized-cut spectral clustering. Após agrupamento, adicionar fator espacial. - Abordagem de crescimento regional - Detecção mudança de padrão: 15 referências bibliográficas (sem nome das técnicas definidos) - Detecção de padrão entre as séries: três referências bibliográficas (sem nome) 	<ul style="list-style-type: none"> - Detectar mudança padrão: Regressão de Joinpoint Ferramenta: Joinpoint Regression Program	
	Multivariada	<ul style="list-style-type: none"> - Equação de estimativa generalizada - Regressão logística - Regressão multinível - Regressão ponderada geograficamente - Testes correlação Ferramentas: SPSS, SpatialStats Software, SAS, pacote stats, lme4 e GWModel do R, GeoDa.	<ul style="list-style-type: none"> - Kriging - Modelos espaciais autoregressivos - Regressão ponderada geograficamente 	<ul style="list-style-type: none"> - Teste qui-quadrado Ferramenta: Python	

Tabela 3.1 – Etapas do estudo e proposta de abordagem técnica de cada etapa.

Durante a pesquisa, contou-se com o fundamental apoio dos professores Barrozo (2023), Saldiva (2023) e Scanavino (2023). A Professora Ligia Barrozo, pertencente ao Departamento de Geografia da FFLCH da USP, é especializada em Geografia Médica e análise espacial, investigando a relação entre urbanização e saúde em São Paulo. O Professor Paulo Saldiva, do Departamento de Patologia da FMUSP, é amplamente reconhecido também por suas pesquisas em Cidades e Saúde Humana. O Professor Marco Scanavino, docente no Departamento de Psiquiatria da USP, tem expertise centrada em temas como comportamento sexual compulsivo e HIV/AIDS, liderando o braço brasileiro da Pesquisa Internacional da Sexualidade. Estes professores foram essenciais ao recomendarem artigos para leitura, apoiarem com vasto conhecimento especializado e sugerirem alterações e oportunidades no estudo, tanto na qualificação do mestrado quanto em reuniões de trabalho.

Para gerenciar o código, optou-se por utilizar o GitHub (2020) na conta Suicities, (2022). Com o GitHub é possível armazenar, rastrear e colaborar em projetos de *software open-source*. É uma plataforma que facilita a colaboração de desenvolvedores, permite controle de versões de modo que não há prejuízo se um usuário comete algum erro no código ou se ele gostaria de personalizar uma parte do código para uso próprio (GITHUB, 2020).

Nas etapas em que se codifica com Python para resolução, utiliza-se o *software* Anaconda (ANACONDA, 2023) para isso. No Anaconda, mais de 250 bibliotecas do Python para uso de ciência de dados já são instaladas automaticamente no ambiente computacional, além de softwares de ambiente de desenvolvimento interativo (*interactive development environment – IDE*) que são idealizadas para cientista de dados, como o Jupyter Notebook (JUPYTER, 2023), IDE utilizada neste projeto. O Jupyter Notebook é uma interface que ilustra o processo de análise de dados passo a passo, consolidando em um só local códigos, imagens, textos, resultados/saídas do código, gráficos e mapas na ordem em que cada etapa foi realizada. Dessa forma, é possível documentar todo o processo de análise didaticamente (JUPYTER, 2023).

Os resultados das análises são expostos nos painéis de visualização do Tableau Public (TABLEAU, 2022), uma ferramenta de visualização de dados que

apresenta uma interface amigável para o usuário, e que foi mais ágil ao tratar dados geográficos que as outras duas ferramentas testadas no projeto (Microsoft Power BI e biblioteca geopandas do Python). O painel final para constante monitoramento do suicídio contempla as seguintes análises e é possível analisar por município, estado ou pelo Brasil em sua totalidade:

- Taxa de suicídio ao longo do tempo, com abertura por gênero masculino/feminino, idade.
- Comparação da taxa de suicídio por idade e gênero masculino/feminino entre dois anos definidos pelo usuário para possibilitar avaliar alguma aplicação de política pública.
- Mapa de calor da taxa de suicídio no Brasil.
- Agrupamentos espaciais.
- Agrupamentos espaço-temporais.
- Tabela com classificação das cidades com maiores tendências temporais de aumento ou queda do suicídio.
- Análise do grau de ruralidade dos municípios considerando os agrupamentos de alto, risco e sem risco da taxa de suicídio.

O painel está disponível no repositório Dada (2022).

A Figura 3-1 ilustra o desenho da arquitetura da solução para viabilizar o estudo, com breves explicações sobre o que é feito em cada etapa. No lado esquerdo da imagem (letra “A” na imagem), encontra-se ilustrada a captura das bases de dados utilizadas no projeto, detalhadas no item 3.1 DADOS. Utiliza-se Python para fazer a transformação e agregação desses dados (letra “B”). Além disso, calcula-se a taxa de suicídio com Python, conforme descrito em 3.2 CÁLCULO DA TAXA DE SUICÍDIO. Após essa etapa, utiliza-se o *software* SatScan, localizado ao centro e acima na imagem (letra “C”), para a análise espaço-temporal apresentada em 3.4 ANÁLISE ESPACIAL, e o *software* Joinpoint, localizado ao centro abaixo na imagem (letra “D”), para análise temporal que é descrita em 3.5 ANÁLISE TEMPORAL. Os resultados de todas as etapas são consolidados no Python, ilustrado na imagem ao lado direito (letra “E”), e é realizada a análise multivariada apresentada em 3.6 ANÁLISE

MULTIVARIADA. Finalmente, expõem-se todos os resultados no Tableau, último item ilustrado na imagem ao lado direito (letra “F”).

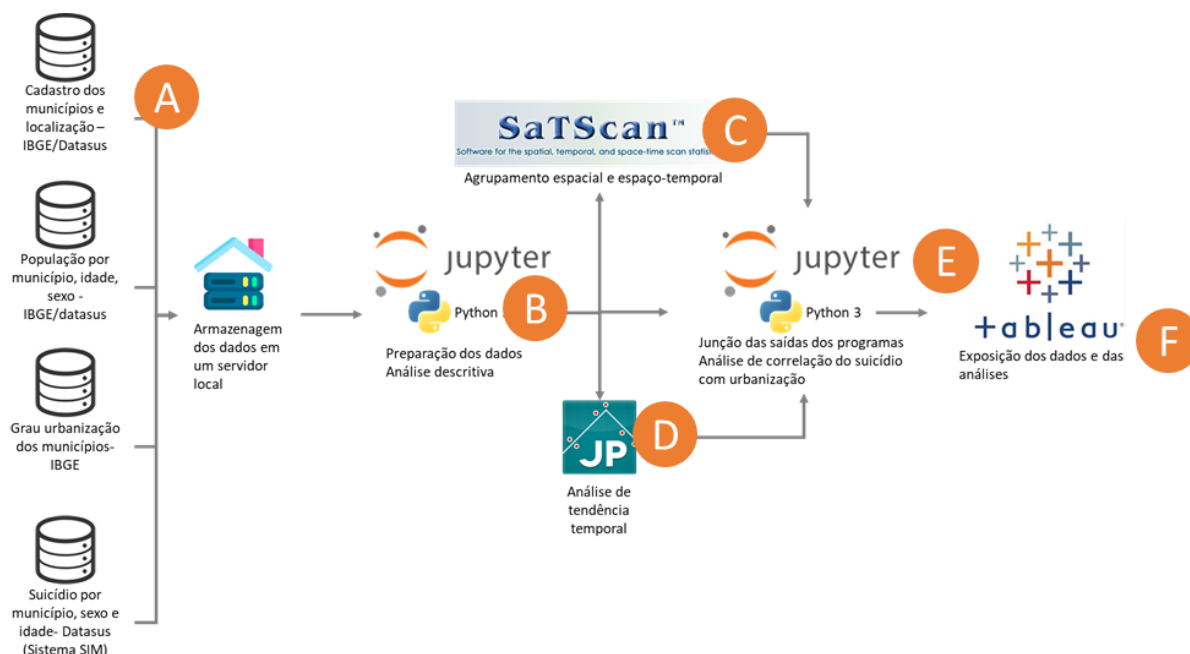


Figura 3-1 – Desenho da arquitetura da solução.

3.1 DADOS

O estudo contempla quatro temas/domínios de dados: (1) cadastro dos municípios com sua latitude e longitude; (2) tamanho da população por município, ano, idade e gênero masculino/feminino; (3) grau de urbanização dos municípios e (4) suicídio por município, gênero masculino/feminino, idade e ano. O período compreendido é de 1996 a 2019. O início da análise é em 1996, por conta da Classificação Internacional de Doenças, versão 10 (CID – 10) que se iniciou nesse mesmo ano. Na época do início deste estudo, os dados disponíveis sobre suicídio abrangiam até o ano de 2019. Contudo, em outubro de 2023, os registros de 2020 e 2021 estavam acessíveis e são incluídos no painel do Tableau.

Para cada domínio de dado, conforme orientada a literatura (EL-SAPPAGH; HENDAWI; EL BASTAWISSY, 2011), é realizado o descobrimento do melhor

repositório de origem, depois captura-se o dado no formato em que ele se encontra. Em seguida, transforma-se a base para o formato desejado (no nosso caso, csv) e valida-se a base para uso. Esse é o processo de extração, transformação e carregamento dos dados, em inglês *Extract, Transform and Load (ETL)*, com a adição de uma etapa de validação.

Essas etapas para o primeiro domínio (cadastro dos municípios) estão ilustradas na Figura 3-2 e são descritas nos próximos parágrafos.

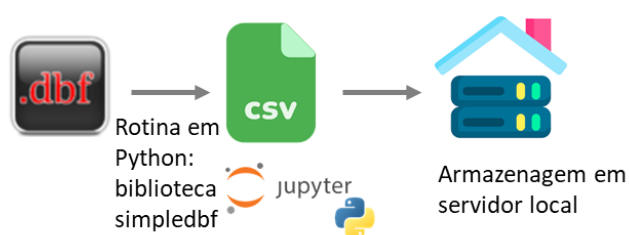
Domínio: cadastro de municípios com latitude e longitude

1 Descobrimiento da fonte:



2 Captura, transformação e carga do dado para consumo:

Transformação arquivo .dbf em .csv



3 Validação:

- Código de município é único e distinto para cada município
- Todos municípios possuem latitude e longitude

Figura 3-2 – Processo de captura das informações de município

Para se obter o cadastro dos municípios e localização geográfica de cada município, utiliza-se o arquivo oferecido pelo Sistema de Informação sobre Mortalidade (SIM) no DATASUS, porque consolida o histórico dos municípios, inclusive os inativos e transferidos e os que foram referenciados pelo SIM em DATASUS, (2022d). Outra base de localidade é a presente no IBGE (2022b).

A base disponibilizada tanto pelo DATASUS quanto pelo IBGE está em formato dbf. Esse formato representa uma base de dados originada por um banco de dados chamado dBase, um dos primeiros sistemas de banco de dados no mundo (DBASE, 2022). A base de dados é transformada para formato csv, pois é um formato de arquivo que é aceito nos softwares que são utilizados para o presente estudo: Jupyter, SatScan e Joinpoint. Para essa transformação, se utiliza a biblioteca `simpledbf` do python no Jupyter.

Na etapa de validação, verifica-se se o código de município provido pela base é único e universal e são excluídos municípios sem latitude e longitude (0,1% da base).

O segundo domínio (tamanho da população por município, gênero masculino/feminino, idade e ano) é obtido pela plataforma do DATASUS (DATASUS, 2022c), que se alimenta de bases criadas no IBGE (IBGE, 2023). A escolha pela origem no DATASUS e não no IBGE diretamente foi porque a primeira plataforma possui todos os dados em uma única base, já a segunda plataforma disponibiliza uma base de dados para cada ano, adicionando uma complexidade no processo pois após a captura das bases, elas teriam que ser empilhadas.

O tamanho da população pode ser calculado por aferimento em pesquisa nacional (Censo) ou por projeção em anos em que não há Censo. Nos anos 2000 e 2010 (anos de Censo), o IBGE utiliza os dados da pesquisa para calcular o tamanho da população por gênero masculino/feminino, idade e município. Para a projeção da população de 1996 a 2012, o IBGE e o Ministério da Saúde / Datasus aplicaram a metodologia presente em Datasus (2022b). Para a projeção para a população de 2013 a 2019, os mesmos órgãos governamentais aplicaram um cálculo de estimativa presente em Datasus (2022a).

O arquivo obtido já é no formato desejado, portanto, não há uma etapa de transformação para esse domínio.

Na etapa de validação, observa-se que há 5567 municípios totais, 593 não possuem população em um ou mais anos. 533 (9% da base) não estão na base de 1996, 53 não estão na base entre 1996 e 2000, 1 não está na base entre 1996 e 2008 e 6 não estão na base entre 1996 e 2012. Esses municípios são excluídos ou tratados como exceção de acordo com cada etapa da análise e cada etapa descreve como trata esse caso.

O processo de captura da informação de população está ilustrado na Figura 3-3.

Domínio: tamanho da população por município, ano e sexo

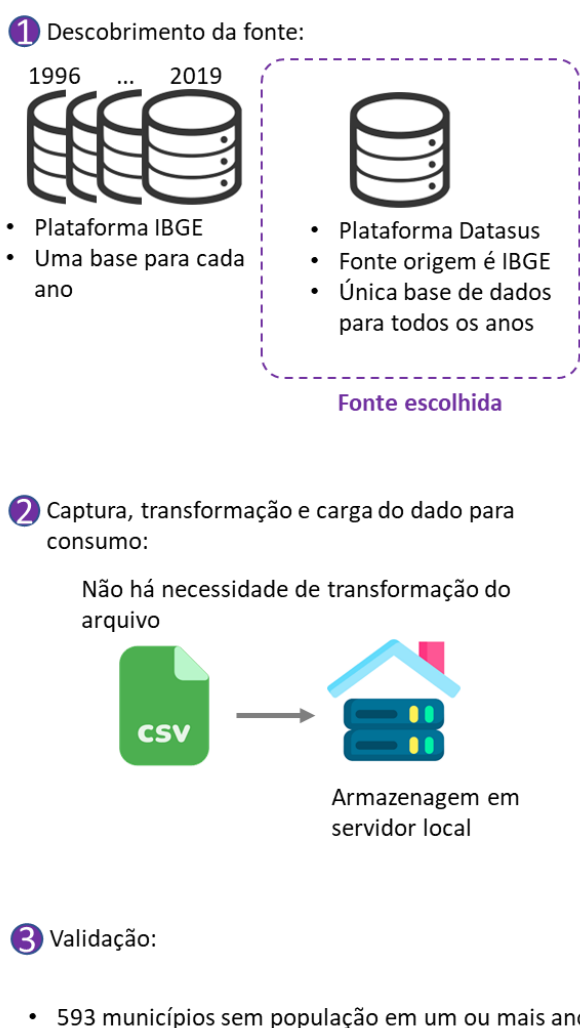


Figura 3-3 – Processo de captura das informações de população

Sobre o terceiro domínio, que é conceito de área urbana ou rural, o Censo considera áreas legalmente definidas como urbana ou rural (IBGE, 2022). O conceito legal está no Código Tributário Nacional, no Art. 32 da Lei 5.172, de 25 de outubro de 1996 (BRASIL, 1996), que diz sobre como decidir se um imposto é de uma área urbana ou rural:

“O imposto, de competência dos Municípios, sobre a propriedade predial e territorial urbana tem como fato gerador a propriedade, o domínio útil ou a posse de bem imóvel por natureza ou por acessão física, como definido na lei civil, localizado na zona urbana do Município.

§ 1º Para os efeitos deste imposto, entende-se como zona urbana a definida em lei municipal; observado o requisito mínimo da existência de melhoramentos indicados em pelo menos 2 (dois) dos incisos seguintes, construídos ou mantidos pelo Poder Público:

I - meio-fio ou calçamento, com canalização de águas pluviais;

II - abastecimento de água;

III - sistema de esgotos sanitários;

IV - rede de iluminação pública, com ou sem posteamento para distribuição domiciliar;

V - escola primária ou posto de saúde a uma distância máxima de 3 (três) quilômetros do imóvel considerado.

§ 2º A lei municipal pode considerar urbanas as áreas urbanizáveis, ou de expansão urbana, constantes de loteamentos aprovados pelos órgãos competentes, destinados à habitação, à indústria ou ao comércio, mesmo que localizados fora das zonas definidas nos termos do parágrafo anterior.”

O IBGE (2017) propôs uma forma complementar de zona rural e urbana. Neste conceito, além da classificação de zona rural e urbana, existem escalas intermediárias de aglomerações, como a área Intermediária adjacente à área urbana. Neste estudo, é utilizado o novo conceito proposto pelo IBGE. As bases estão em IBGE (2022a).

O arquivo obtido já é no formato desejado, portanto, não há uma etapa de transformação para esse domínio. Na etapa de validação, há 0,2% dos municípios

sem a classificação de ruralidade. Para os municípios que não possuíam a categorização de ruralidade, realizou-se uma pesquisa manual no *Google* para cada cidade, classificando-as conforme os critérios estabelecidos pelo IBGE.

A Figura 3-4 ilustra como foi o processo de captura dessa informação.

Domínio: grau de urbanização dos municípios

1 Descobrimto da fonte:



Fonte escolhida

2 Captura, transformação e carga do dado para consumo:

Não há necessidade de transformação do arquivo



3 Validação:

- 0,02% dos municípios sem informação de grau de urbanização/ruralidade.

Figura 3-4 – Processo de captura das informações de urbanização

Por fim, o quarto domínio contempla os suicídios dos municípios do Brasil por dia, mês e ano do óbito, gênero masculino/feminino e idade, com os dados do período de 1996 a 2019.

Devido à complexidade inerente à coleta e verificação dos dados de mortalidade, conforme discutido na seção “2.2.1 – DATASUS”, as bases de dados de mortalidade apresentam uma defasagem temporal de um ano ou mais. O presente estudo iniciou em 2021 e abrange informações até o ano de 2019. Em outubro de 2023, os registros de 2020 e 2021 estavam acessíveis e são incluídos no painel do Tableau.

A base de suicídio é obtida a partir do Tabnet, na fonte (FTP DATASUS, 2020). Nessa fonte, os dados vitais estão em arquivos em formato dbf segregados por unidade federativa e ano, totalizando 621 arquivos. Fez-se uma rotina em Python para extração, transformação e carga (*Extract, Transform and Load* - ETL) dessas bases de dados. Os dados obtidos são empilhados em uma única base de dados utilizando a biblioteca pandas do python. O resultado é validado validados pela comparação do número de suicídio por município obtido com a base de dados e com a plataforma de consulta de dados do Tabnet (ferramenta de *Business Intelligence* – BI).

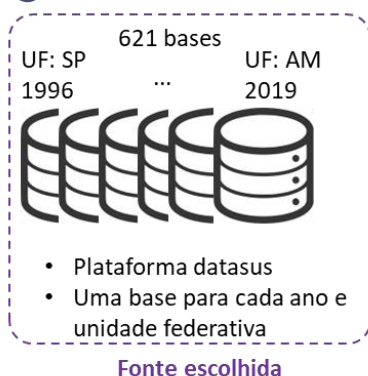
Para o suicídio, são considerados os códigos do CID X60 ao X84, que significam “Lesões autoprovocadas intencionalmente” e o CID Y87, que significa “Sequelas de uma lesão autoprovocada intencionalmente, de agressão ou de um fato cuja intenção é indeterminada”.

Durante a obtenção dos dados, é plausível que haja subnotificação de casos de suicídio devido a fatores de natureza religiosa, social, cultural, política e econômica, como observado por PALMA, SANTOS e IGNOTTI (2020). O presente estudo não aborda explicitamente a influência da subnotificação em seus resultados. Uma perspectiva de investigação futura consiste na aplicação de técnicas de ciência de dados para avaliar a presença de subnotificação de casos de suicídio, considerando a análise da magnitude das causas de morte em relação a fatores religiosos e sociodemográficos específicos em diferentes localidades.

A Figura 3-5 ilustra como foi o processo de captura das informações de suicídio:

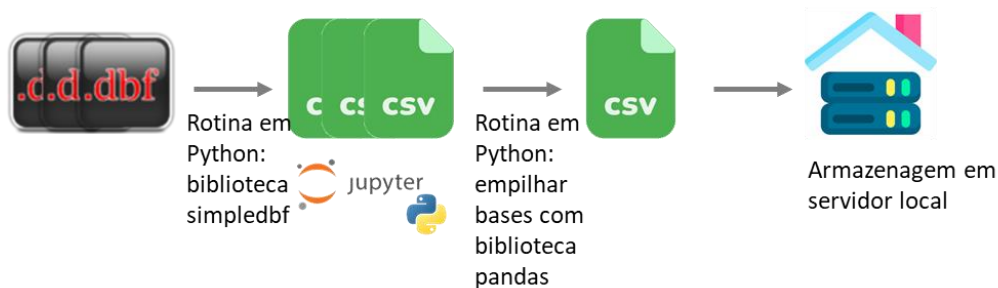
Domínio: suicídios

1 Descobrimento da fonte:



2 Captura, transformação e carga do dado para consumo:

Transformação arquivo .dbf em .csv e depois unifica os .csv em um único .csv



3 Validação:

- Dados obtidos são validados com a plataforma de BI do Datasus (Tabnet). Se ambos resultam em um mesmo número de suicídio, a captura foi realizada com êxito.

Figura 3-5 – Processo de captura das informações de suicídio

Os quatro domínios de informação se correlacionam segundo o modelo lógico relacional apresentado na Figura 3-6.

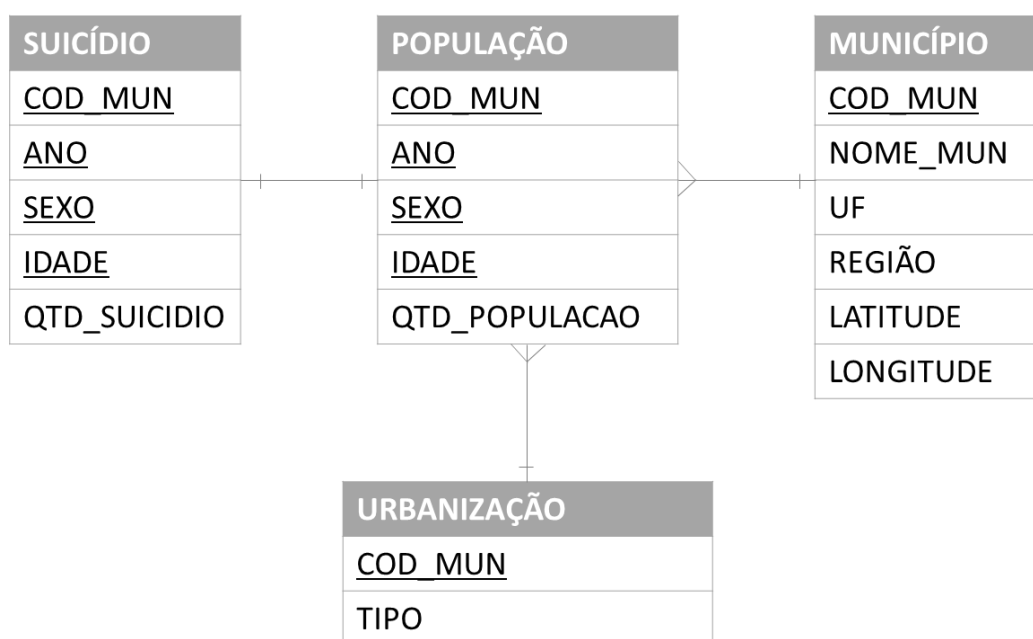


Figura 3-6 – Modelo lógico relacional das bases necessárias para o estudo

Para se realizar os estudos, as tabelas são condensadas em uma única base de dados.

3.2 CÁLCULO DA TAXA DE SUICÍDIO

Existem três formas do cálculo da taxa de suicídio: **(1)** taxa ajustada por idade (padronização direta), conforme a OMS recomenda, **(2)** taxa ajustada por idade e gênero masculino/feminino e **(3)** a taxa calculada por modelos bayesianos para suavização espacial. Os resultados podem variar por conta dessas diferentes formas de se calcular. Além da diferenciação do cálculo da taxa, ainda há a diferenciação dos métodos de cálculo das projeções de tamanho da população. Em anos em que não há Censo, é feita uma projeção da população, o que pode afetar os resultados.

Nesta pesquisa, não foram calculados os impactos dos diferentes métodos de cálculo da projeção da população.

A taxa ajustada por idade foi calculada baseada na distribuição sugerida pela OMS (AHMAD; BOSCHI-PINTO; LOPEZ, 2001) para comparabilidade da taxa de suicídio brasileira com a de outros países.

Para todas as outras análises do estudo, foi utilizada a taxa de suicídio ajustada por gênero masculino/feminino e idade. A taxa de suicídio pode apresentar mudança de padrão a partir de 2011, porque foi nesse ano que passou a ser obrigatória a notificação do ato (PALMA; SANTOS; IGNOTTI, 2020).

As faixas de idade escolhidas são iguais às que a OMS utiliza em suas análises de mortalidade e suicídio, presente em WHO, (2022) e em WHO, (1999). São elas: 0 a 4 anos, 5 a 14 anos, 15 a 24 anos, 25 a 34 anos, 35 a 44 anos, 45 a 54 anos, 55 a 64 anos, 65 a 74 anos e 75 anos e mais.

O cálculo da taxa ajustada é apresentado na equação 4:

$$taxa\ ajustada = \sum_{i=faixa\ inicial\ de\ idade}^{Todas\ as\ faixas\ de\ idade} \left(\frac{\# Suicídios\ na\ faixa\ i}{População\ da\ faixa\ i} \right) \times Proporção\ WHO\ da\ faixa\ i \quad (4)$$

3.3 VALIDAÇÃO DOS DADOS E ANÁLISE DESCRITIVA

Na etapa de validação dos dados da taxa de suicídio, como o método de cálculo da taxa de suicídio é a taxa ajustada e não a taxa suavizada, ocorrem anomalias de valores altos na taxa de suicídio em cidades com o número pequeno de habitantes, além de uma intermitência na taxa de suicídio ao longo dos anos para estes municípios. Isso significa, por exemplo, que uma cidade pequena, com 1.000 habitantes, onde ocorreu um suicídio no ano X e depois não apresentou mais nenhum evento do tipo, pode apresentar a taxa de 100 suicídios por 100.000 habitantes no ano X e nos anos subsequentes ela apresenta zero suicídio, gerando uma inconstância nas análises temporais e um falso alarme de alto risco nos agrupamentos espaciais no ano X.

Para se tratar casos assim, primeiramente as curvas temporais de suicídio de cada município são agrupadas pelo método *Kmeans* no Python, para se verificar qual é o agrupamento que está muito longe do centroide e quais são as tendências temporais dos municípios, buscando entender se há algum agrupamento que apresenta o efeito intermitente, para ser excluído em análises futuras. É avaliado se há um efeito cotovelo (*elbow*) para um determinado número de agrupamentos e se é possível considerar que os agrupamentos convergiram. O *elbow* analisa a distância

quadrada de cada ponto até seu centroide. O número de *clusters* ideal é aquele que minimiza a distância quadrada de cada ponto até seu centroide.

Em um agrupamento em que os valores convergem, normalmente há a indicação de um ponto de inflexão do gráfico, conforme a figura 3.2.

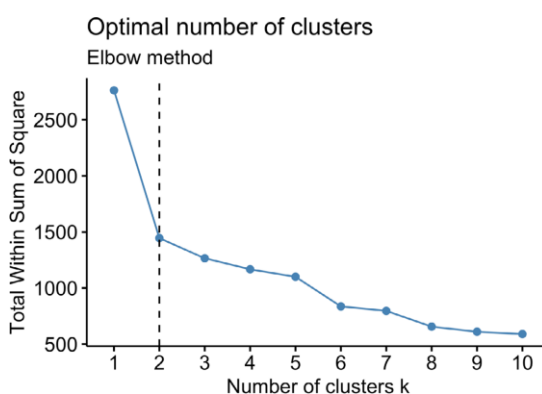


Figura 3-7 – Número de *clusters* ideal igual a 2 analisando o *elbow*.
Fonte: LOPES e GOSLING (2021)

Caso não ocorra o *elbow* no agrupamento de cidades, com indicação de cidades que apresentam intermitência, são excluídos municípios das análises considerando frequência e significância. São consideradas apenas cidades com suicídio em ao menos 75% dos anos para se ter ao menos 18 anos de análise e se poder afirmar se há uma tendência histórica ou se acontece algum evento atípico na região. Ou seja, cidades que registraram suicídio em 18 anos, do total de 24 anos analisados. Assim, 1.214 municípios são considerados na análise, que são 21% do total de 5.659 municípios. Além disso, para melhor análise gráfica do mapa das taxas altas de suicídio, para taxas de suicídio maiores que 13, foi delimitado como 13 a taxa de suicídio. Assim se impede que as taxas mais altas distorçam a análise gráfica, possibilitando-se ver regiões com crescimento de taxa ao longo do tempo.

O número 13 foi escolhido baseado no artigo de Khana et. al (2018), que indica como significativas apenas localizações geográficas que apresentam acima de 20 suicídios em um ano. Essa é uma das técnicas para se tratar *outliers*. Na base de suicídios (FTP DATASUS, 2020)., as cidades brasileiras que apresentam no mínimo 20 suicídios ao ano em número absoluto, seguindo a significância mínima de eventos

em uma região para análise sugerida por Khana et. al (2018), apresentam taxa máxima de suicídio de 13. Então, essa seria a taxa mais confiável e significativa para se considerar nas análises. Taxas mais altas de suicídio tendem a ser de localidades com o número de habitantes pequeno e com o incidente do suicídio em situações anômalas.

Na análise descritiva, como citado anteriormente, investiga-se, por meio de gráficos, tabelas e mapas, uma visão preliminar das cinco áreas de questionamento que direcionam o estudo.

Análises históricas de tendência no Brasil são analisadas, com aprofundamento nas faixas de idade e gênero masculino/feminino para se entender qual é o grupo que está causando o aumento da taxa no Brasil.

Para se verificar, preliminarmente, como o suicídio se altera no espaço ao longo do tempo, mapas da taxa de suicídio no Brasil são feitos considerando um conjunto de anos. Por exemplo, é feita a comparação do mapa elaborado em 2010 e em 2019.

Os *softwares* Excel, Jupyter (Python) e Tableau foram utilizados nessa etapa.

Para elaboração de mapas mais detalhados, são utilizados os pacotes de camadas oferecidos pelo IBGE, 2022c. Utiliza-se a camada *bcim*, que oferece a escala 1: 1.000.000 e a camada utilizada foi a *LIM_Municipio_A* (IBGE - DIRETORIA DE GEOCIÊNCIAS, 2016). O mapa foi elaborado com a biblioteca *Geopandas* do *Python*.

3.4 ANÁLISE ESPACIAL

Com a análise espacial, busca-se aprofundar no pilar: “Quais agrupamentos espaciais podem estar relacionados com os dados sobre suicídio? Estes agrupamentos mudam ao longo do tempo? Há algum agrupamento de maior relevância? Há agrupamentos de baixo risco para destacar?”.

Para identificar possíveis agrupamentos espaciais, é feito o agrupamento proposto por Kulldorff (1997) da taxa de suicídio mais recente. Dessa forma, é possível identificar agrupamentos de alto risco para alertar os governos locais sobre a possibilidade de um risco de alerta.

É feita a análise espaço-temporal proposta também por Kulldorff (1997) para verificar se estes poderiam se alterar ao longo do tempo.

Para preparar os dados para o SatScan, ferramenta que faz o agrupamento especial proposto por Kulldorff, primeiro é necessário calcular a taxa de suicídio padronizada indiretamente, ajustada por idade e gênero masculino/feminino. Na padronização indireta, são comparados os casos observados e os esperados de cada unidade que está sendo analisada. Os casos esperados são calculados a partir da taxa do evento que se observa em toda a população estratificada, multiplicada pela população estratificada existente naquela unidade que está sendo observada (TRIPEPI et al., 2010). Por exemplo, o número esperado de suicídios na cidade de São Paulo em homens de 15 a 24 anos é a população de homens de 15 a 24 anos, que moram em São Paulo, multiplicada pela taxa de suicídio geral que acontece no Brasil em homens dessa faixa etária.

O cálculo de suicídios esperado foi feito estratificando gênero masculino/feminino e faixa de idade. Assim, foram excluídos os casos de suicídio em que não foi reportado gênero masculino/feminino ou idade. Para rodar a análise, determinou-se que a distribuição dos casos é de Poisson, o tamanho máximo de um agrupamento foi escolhido em 15% da população, foram escolhidas 999 replicações de Monte Carlo para se fazer o teste, e para o estudo foi determinado que não poderia haver agrupamentos se sobrepondo.

Para se entender como os agrupamentos espaciais mudam ao longo do tempo, deseja-se criar agrupamentos espaço-temporais no SatScan. Para isso, primeiramente se analisou quais cidades não existiam dos anos de 1996 a 2019. Dos 5567 municípios totais, 593 não aparecem em um ou mais anos. 533 (9% da base) não estão na base de 1996, 53 não estão na base entre 1996 e 2000, 1 não está na base entre 1996 e 2008 e 6 não estão na base entre 1996 e 2012.

Sendo assim, entende-se que a análise espaço-temporal deva considerar o período de 1997 a 2019, retirando-se os municípios da análise que não aparecem em um dos anos. No total foram 60 municípios excluídos (53 que não estão entre 1996 e 2000, 1 de 1996 a 2008 e 6 de 1996 a 2012). Os municípios excluídos foram:

- Região Norte: Mojuí dos Campos (Pará)
- Região Nordeste: Nazária (Piauí), Pau D'Arco do Piauí (Piauí), Jundiá (Rio Grande do Norte), Jequiá da Praia (Alagoas), Barrocas (Bahia), Luís Eduardo Magalhães (Bahia).
- Região Sudeste: Governador Lindenberg (Espírito Santo), Mesquita (Rio de Janeiro).
- Região Sul: Balneário Rincão (Santa Catarina), Pescaria Brava (Santa Catarina), Aceguá, Almirante Tamandaré do Sul, Arroio do Padre, Boa Vista do Cadeado, Boa Vista do Incra, Bozano, Canudos do Vale, Capão Bonito do Sul, Capão do Cipó, Coqueiro Baixo, Coronel Pilar, Cruzaltense, Forquetinha, Itati, Jacuizinho, Lagoa Bonita do Sul, Mato Queimado, Novo Xingu, Paulo Bento, Pedras Altas, Pinhal da Serra, Pinto Bandeira, Quatro Irmãos, Rolador, Santa Cecília do Sul, Santa Margarida do Sul, São José do Sul, São Pedro das Missões, Tio Hugo, Westfalia. Todas as cidades, com exceção de Balneário Rincão e Pescaria Brava, são do estado do Rio Grande do Sul.
- Centro-Oeste: Paraíso das Águas (Mato Grosso do Sul), Bom Jesus do Araguaia (Mato Grosso), Colniza (Mato Grosso), Conquista D'Oeste (Mato Grosso), Curvelândia (Mato Grosso), Nova Nazaré (Mato Grosso), Nova Santa Helena (Mato Grosso), Novo Santo Antônio (Mato Grosso), Rondolândia (Mato Grosso), Santa Cruz do Xingu (Mato Grosso), Santa Rita do Trivelato (Mato Grosso), Santo Antônio do Leste (Mato Grosso), Serra Nova Dourada (Mato Grosso), Vale de São Domingos (Mato Grosso), Campo Limpo de Goiás (Goiás), Gameleira de Goiás (Goiás), Ipiranga de Goiás (Goiás), Lagoa Santa (Goiás).

3.5 ANÁLISE TEMPORAL

Com a análise temporal, buscou-se aprofundar nos questionamentos: “Historicamente, quais são as cidades com maiores crescimentos da taxa e as com maiores quedas? Quais municípios e quais grupos de gênero masculino/feminino e idade contribuem para o aumento da taxa de suicídio no Brasil?”.

Para o questionamento, foi calculada a regressão de Joinpoint (NCI, 2020) para a taxa anual de suicídio de cada combinação de gênero masculino/feminino e faixa de idade.

Foi utilizado o programa Regressão de Joinpoint (NCI - SRP, 2022) e o *software* Jupyter (Python).

A taxa ajustada foi calculada no Python, considerando-se que para o suicídio não há heterocedasticidade conforme adotado na literatura (EXETER e BOYLE, 2007) (MESOUDI, 2009) (JONES, 2013), ou seja, a variância dos erros da taxa de suicídio ajustada é constante, não se altera ao longo dos anos. conforme ilustrado na figura a Figura 3-8. Como parâmetros adicionais para a análise, utilizou-se *grid-search* para descobrir a melhor localização dos intervalos.

O número máximo de quebras/cortes escolhido foi 4, parâmetro recomendado considerando o total de pontos que a série temporal possui. O National Cancer Institute (2022c), recomenda que no mínimo deve haver 7 pontos para que haja uma quebra/corte, e ao menos dois pontos entre cada quebra/corte. Para seleção de qual o número de quebras/cortes é adotado, utiliza-se a permutação com amostragem Monte Carlo e 4.499 permutações. Para a análise temporal, foram considerados 5.507 municípios, como foi feito no agrupamento espaço-temporal do Satscan.

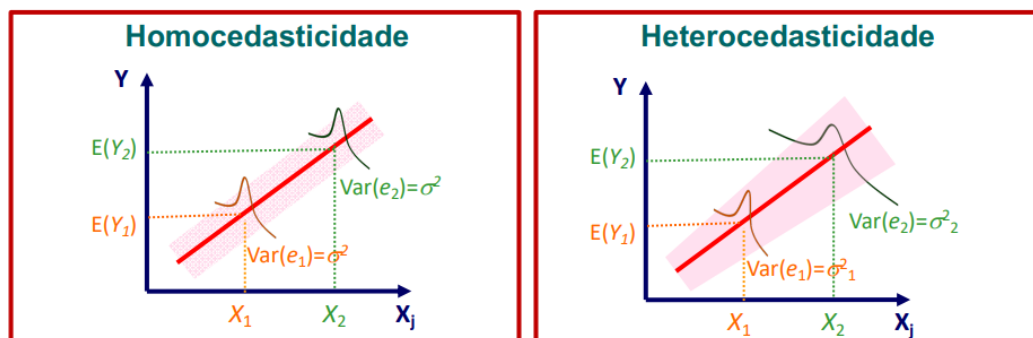


Figura 3-8 – Diferença entre Homocedasticidade (à esquerda, com a variância constante ao longo do tempo) e Heterocedasticidade (à direita, com a variância aumentando ao longo do tempo).
Fonte: MAIA (2017)

Para se evitar casos em que o programa indica que há uma tendência histórica, mas não há tendência visualmente, conforme mencionado no Item “2.3.2.5 – Análise Temporal”, foi priorizada a redução dos casos de falso positivo. Assim, a pesquisa levantou as cidades que precisam, certamente, de uma ação preventiva, ou aquelas que, certamente, são referência no combate ao suicídio, apontadas no subitem 4.3 - ANÁLISE TEMPORAL

Para ilustrar o parágrafo anterior, as figuras desde Figura 3-9 até a Figura 3-13 ilustram cidades em que a saída da regressão de Joinpoint foi significativa, ou seja, o software indica que há uma tendência histórica na taxa de suicídio daquele município, mas, visualmente, não é possível afirmar que há tendência. As cidades foram selecionadas aleatoriamente nas quais o *software* indica uma tendência estatisticamente significativa de aumento da taxa, contudo, observa-se um desvio considerável na regressão. Cada figura representa a taxa de suicídio histórica de um município. Observa-se a seguir que não é possível concluir que há tendência de aumento da taxa de suicídio em cada uma das cidades apresentadas, conforme a regressão de Joinpoint aponta, sugerindo uma possível inconsistência no método. Por isso, utilizam-se critérios mais conservadores para afirmar que há tendência histórica.

O município de Silves, ilustrado na Figura 3-9, é indicado como um município com tendência de crescimento da taxa de suicídio, porém somente no ano de 2019 a taxa foi maior que zero. 2019 pode ser um ano atípico e não uma tendência, conforme o *software* aponta.

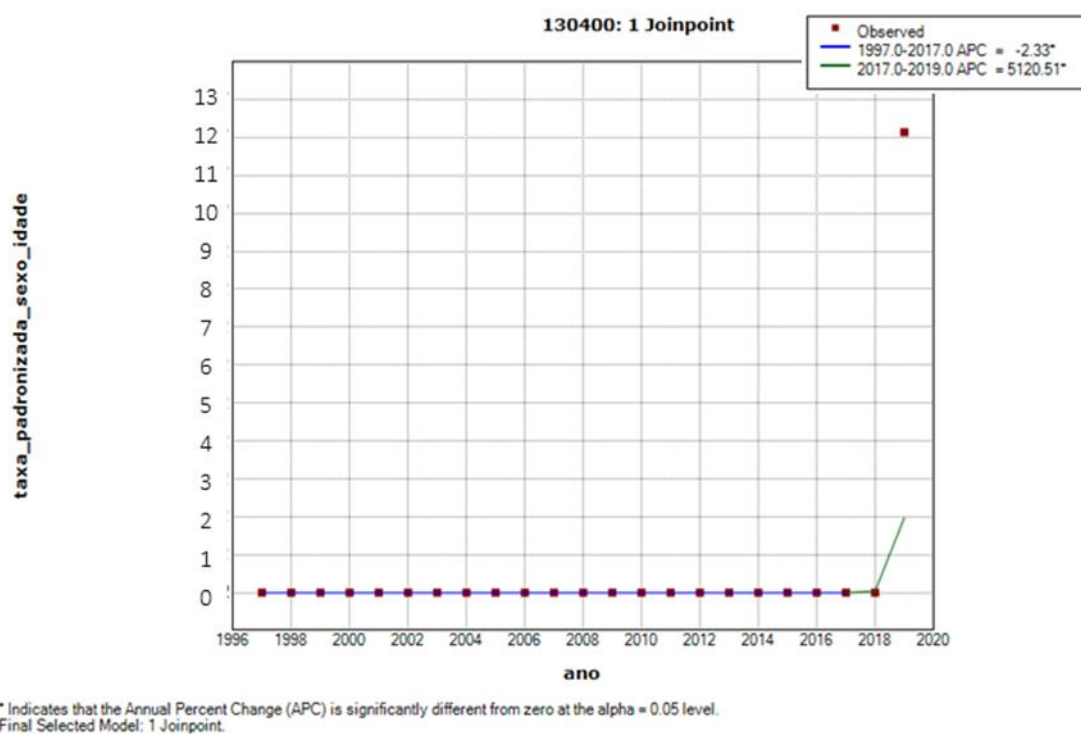


Figura 3-9 – Histórico da taxa de suicídio no município Silves (código município 130400).

O município de Tijucas do Sul, ilustrado na Figura 3-10, é indicado como um município com tendência de crescimento exponencial da taxa de 1996 a 2014, porém se observa no gráfico uma tendência de estabilidade entre 0 ou 10 na taxa de suicídio ao longo dos anos.

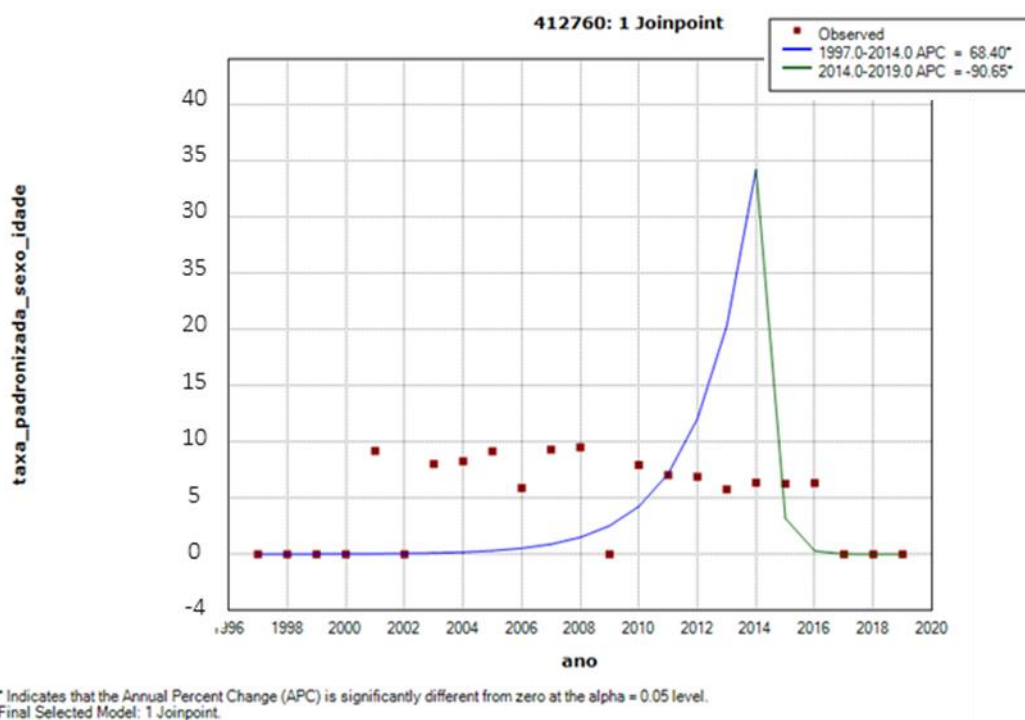


Figura 3-10 – Histórico da taxa de suicídio no município Tijucas do Sul (código do município 412760).

O município de Cássia dos Coqueiros, ilustrado na Figura 3-11, pela análise de Joinpoint, indica que há uma tendência de crescimento da taxa desde 1999 (APC de 48.43 significativamente diferente de zero) e, observando o gráfico, se observa um aumento da taxa somente em 2014 a 2016.

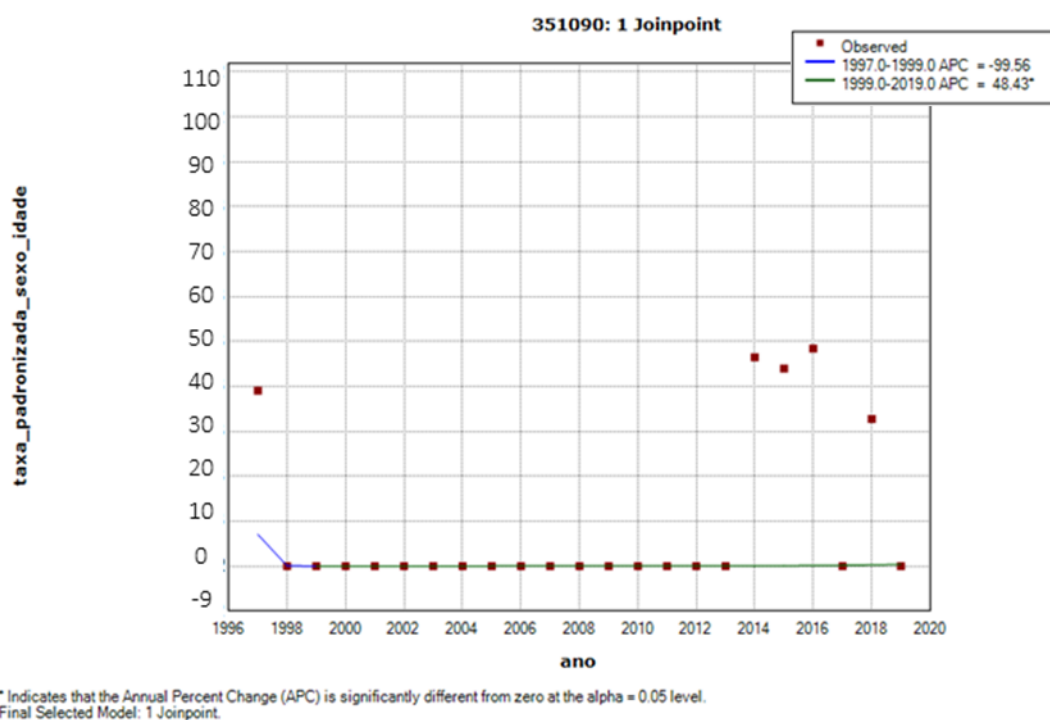


Figura 3-11 – Histórico da taxa de suicídio no município Cássia dos Coqueiros (código do município 351090).

O município de Santa Inês, ilustrado na Figura 3-12, apresenta pela análise de Joinpoint uma tendência histórica desde 2017 de crescimento (APC significativo de 8933), porém apenas em 2019 se observa suicídio no município.

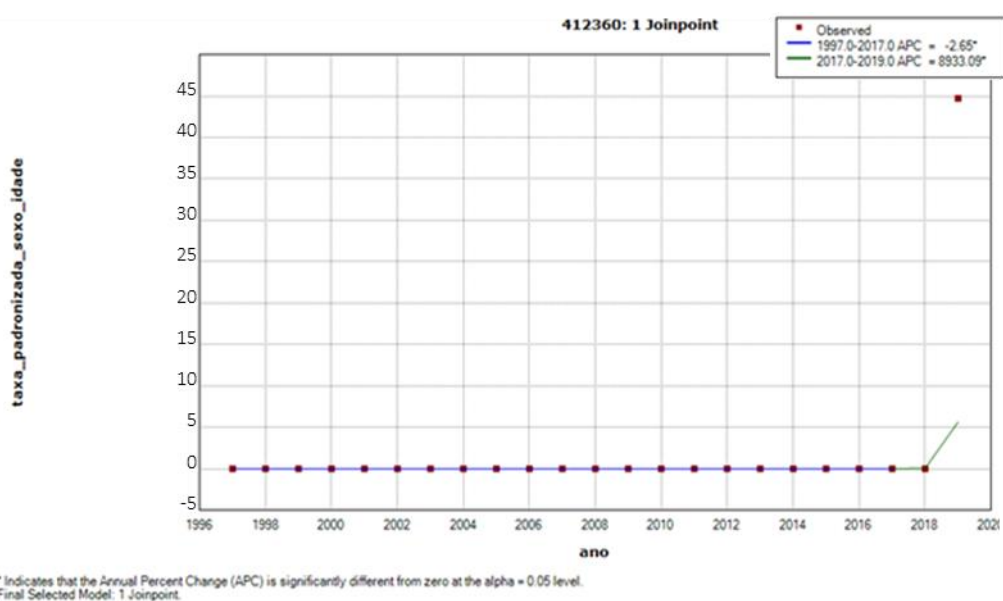


Figura 3-12 – Histórico da taxa de suicídio no município Santa Inês (código do município 4123600).

No município de Água Santana, ilustrado na Figura 3-13, a regressão de Joinpoint indica que há tendência de queda do suicídio nos últimos anos (APC de 2017 a 2019 é -99 significativo), mas a cidade mais aparenta apresentar uma anomalia nos anos de 2016 e 2017.

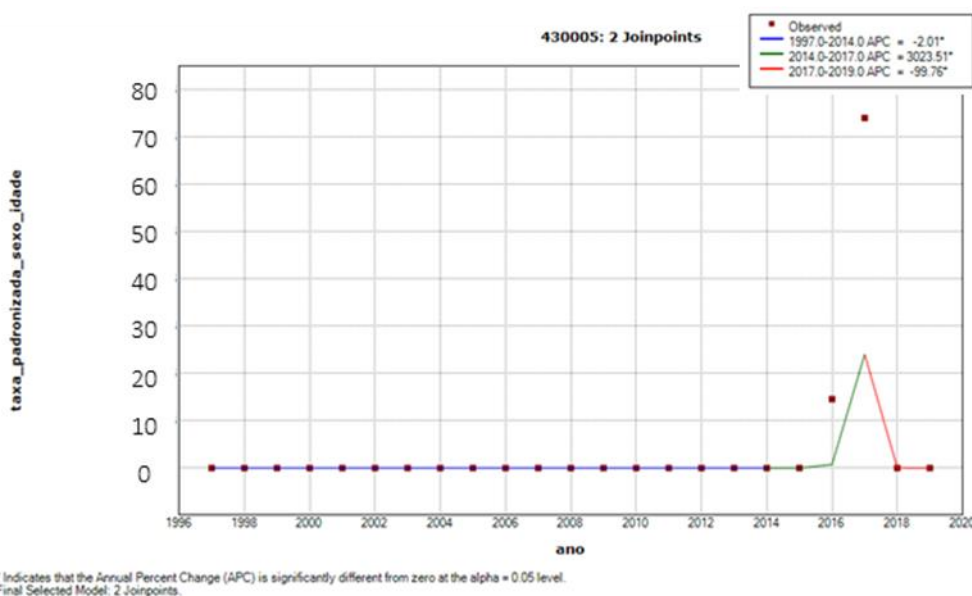


Figura 3-13- Histórico da taxa de suicídio no município Água Santana (código do município 430005)

Para se considerar que há uma tendência, além do p-valor abaixo de 5%, foi considerada a amplitude do APC, levando em conta seu intervalo de confiança. Sabe-se que o intervalo de confiança considera o erro obtido da regressão. Então, uma amplitude alta significa que o modelo pode errar bastante em pontos percentuais. Ou seja, uma alta amplitude de APC significa também um alto erro quadrático médio da regressão, conforme a Figura 3-14.

Analisando as distribuições da amplitude visualmente no gráfico, além do APC significativamente diferente de zero pelo teste “t”, considera-se um erro quadrático médio máximo de 3,2, que é o ponto de inflexão da curva (elipse vermelha na figura), ou uma amplitude do intervalo de confiança de até 26 no APC. Além disso, para as

regressões que apresentam pontos de inflexão, ou seja, *joinpoints*, desconsidera-se APC acima de 90, porque isso significa que são cidades que sempre apresentaram taxa zero de suicídio, e nos últimos 2 ou 3 anos apresentaram a taxa, mas ainda não se pode afirmar que é uma tendência histórica.

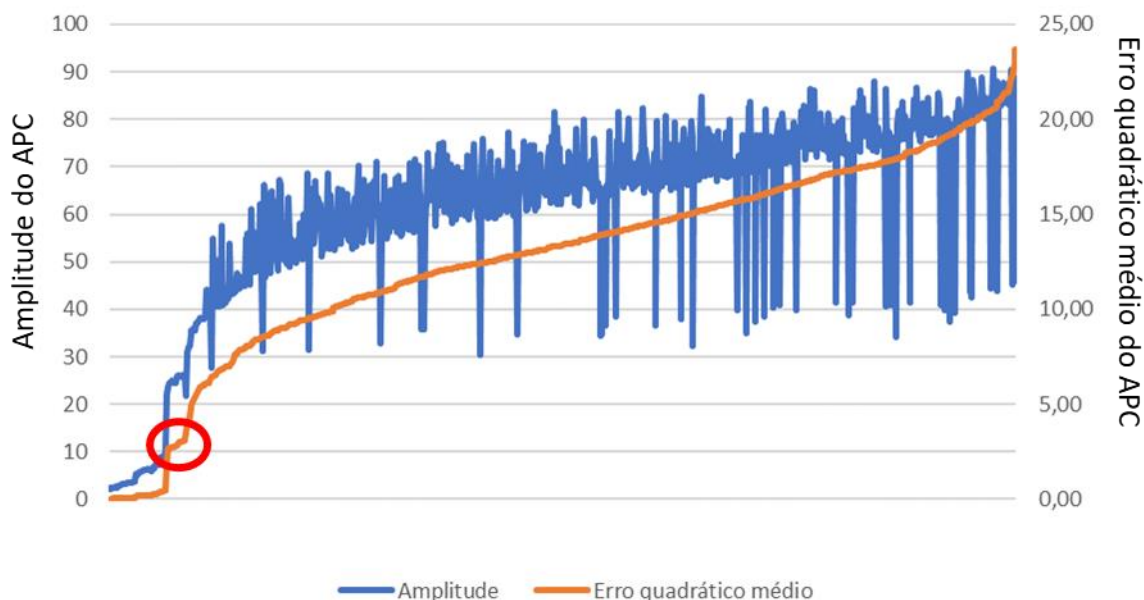


Figura 3-14 - Relação entre o erro quadrático médio (MSE) e a amplitude de APC significativamente diferente de zero. Elipse vermelha aponta ponto de inflexão considerado para as análises.

3.6 ANÁLISE MULTIVARIADA

Com a análise multivariada, deseja-se aprofundar nos pilares: “O percentual de urbanização contribui para explicar as variações das taxas de suicídio?”.

A fim de investigar a existência de uma relação entre as taxas de suicídio e o nível de urbanização de determinada região, foi conduzida uma análise utilizando o teste de qui-quadrado. O teste de qui-quadrado é uma técnica estatística que avalia se há uma associação significativa entre duas variáveis categóricas, comparando a frequência observada de dados em categorias específicas com o que se esperaria se não houvesse relação entre as variáveis.

Neste estudo, esse teste foi aplicado considerando (1) as proporções derivadas do conceito de ruralidade, conforme estabelecido pelo Censo Demográfico

e (2) o agrupamento espacial obtido por Kulldorff dos agrupamentos de "alto risco", "baixo risco" e "sem risco". O objetivo é compreender se regiões mais urbanizadas estão associadas a uma maior prevalência de áreas com alto risco de suicídio, ou, inversamente, se apresentam uma predominância de áreas com baixo risco de suicídio. É possível, contudo, que inicialmente não se observe uma relação direta entre essas variáveis.

A relevância estatística da correlação é confirmada se o p-valor resultante do teste for inferior a 5%. Ou seja, aplicando o teste estatístico, um p-valor inferior a 5% indica uma correlação entre o grau de urbanização de um município e sua tendência de apresentar taxas elevadas ou reduzidas de suicídio. Já um p-valor superior a 5% sugere que não há evidência suficiente para estabelecer uma relação com base nesse teste. Para conduzir a execução do teste utilizou-se o *software* Jupyter (Python) e Excel para visualização do gráfico.

4 RESULTADOS

Apresentam-se, neste Capítulo, os resultados obtidos em cada etapa da metodologia de análise proposta utilizando *data-mining*. Juntos, os resultados contribuem com uma visão holística do histórico do suicídio nas cidades brasileiras. Posteriormente, são apresentadas as conclusões consolidadas da pesquisa e sugestões para pesquisas futuras.

4.1 CÁLCULO DA TAXA DE SUICÍDIO E ANÁLISE DESCRITIVA

A taxa ajustada por idade, obtida pelos dados mais individualizados (menos agrupados) em comparação ao obtido pela OMS (2021), é similar até 2013. Após 2013, a taxa apresentada pela OMS é maior que a calculada neste estudo. A hipótese para essa diferença é que o cálculo de estimativa de população se altera justamente em 2013. A diferença não está no CID, pois é o mesmo usado pela OMS (2018), e também a discrepância não aponta estar no número de suicídios, uma vez que não que a fonte origem geradora dos dados de suicídio é a mesma em todos os anos analisados. O tamanho da população foi retirado do Censo.

Mesmo com as taxas ligeiramente diferentes, há a tendência de crescimento da taxa de suicídio no Brasil, conforme Figura 4-1.

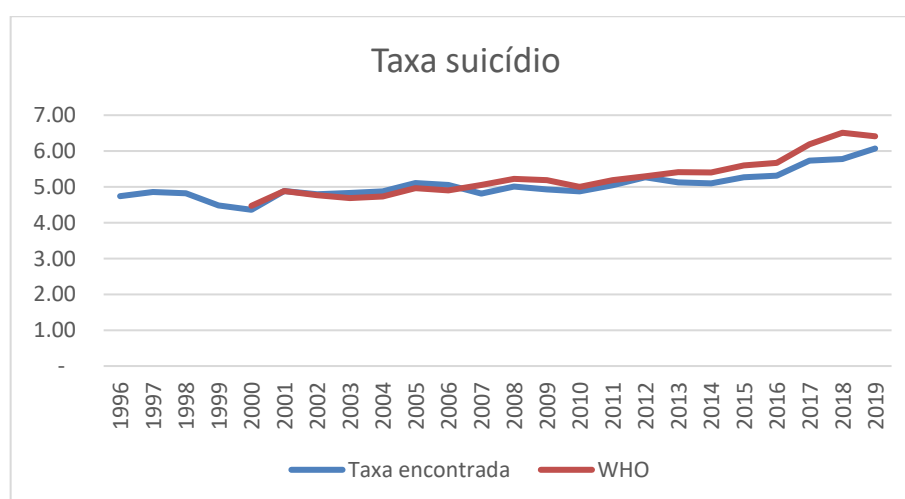


Figura 4-1 – Taxa de suicídio no Brasil calculada neste estudo e o publicado pela OMS.

Já é conhecido que a taxa de suicídio é influenciada pelo gênero. Por isso, foi analisada a taxa de suicídio padronizada segregada pelo gênero feminino/masculino. A taxa de suicídio do gênero masculino é cerca de 3,5 vezes maior que a do gênero feminino. Pode-se concluir que tanto para homens quanto para mulheres, a taxa de suicídio vem crescendo, conforme Figura 4-2.

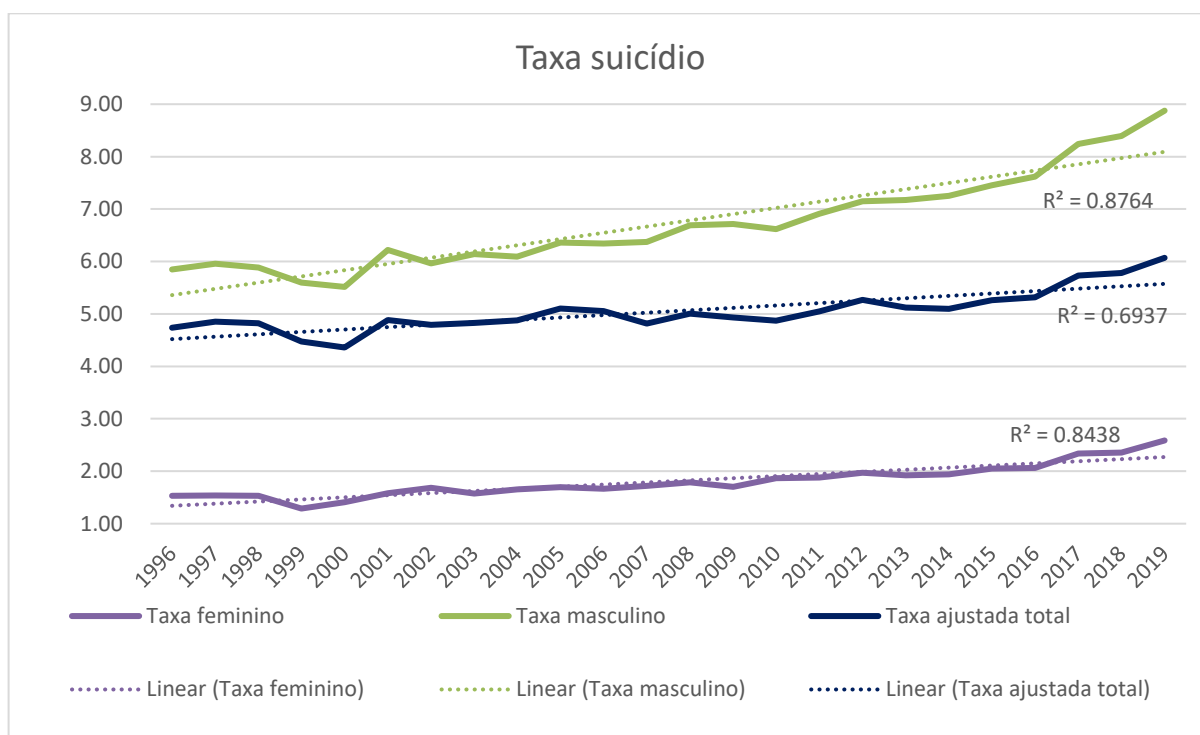


Figura 4-2 – Taxa de suicídio no Brasil por gênero feminino/masculino.

Quando se analisa por idade, observa-se que, conforme mencionado na bibliografia, a taxa de suicídio é maior para faixas maiores de idade e a taxa vem crescendo mais aceleradamente nas faixas mais jovens, principalmente de 15 a 34 anos. Ao longo do período todo, observa-se que apenas a taxa de suicídio acima de 75 anos se mantém estável (Figura 4-3).

Taxa por idade por 100.000 habitantes

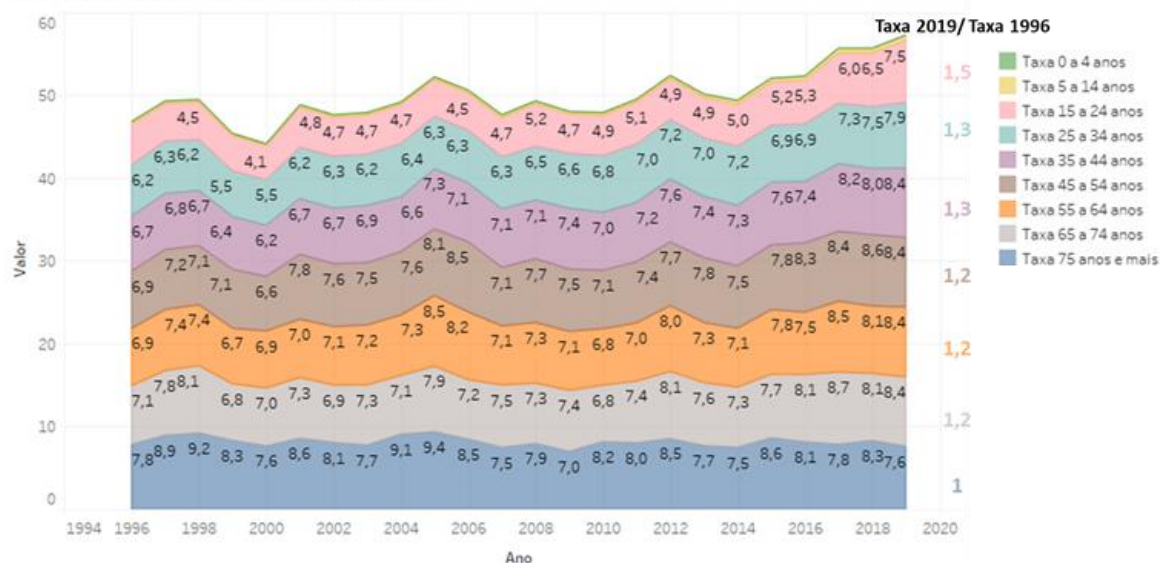


Figura 4-3 – Taxa de suicídio em cada faixa de idade por 100.00 habitantes.

Quando se aprofunda a análise por gênero masculino/feminino e faixa de idade, para se entender qual o grupo que mais cresceu na taxa, conforme a Figura 4-4, verifica-se que tanto no gênero feminino quanto no masculino a taxa apresenta maior aumento nas faixas mais jovens de idade. Em algumas faixas, há uma aceleração maior do crescimento da taxa para o gênero feminino. Porém, quando se analisa o número total de suicídios, o gênero feminino influencia pouco no número total do Brasil porque este gênero apresenta menos casos.

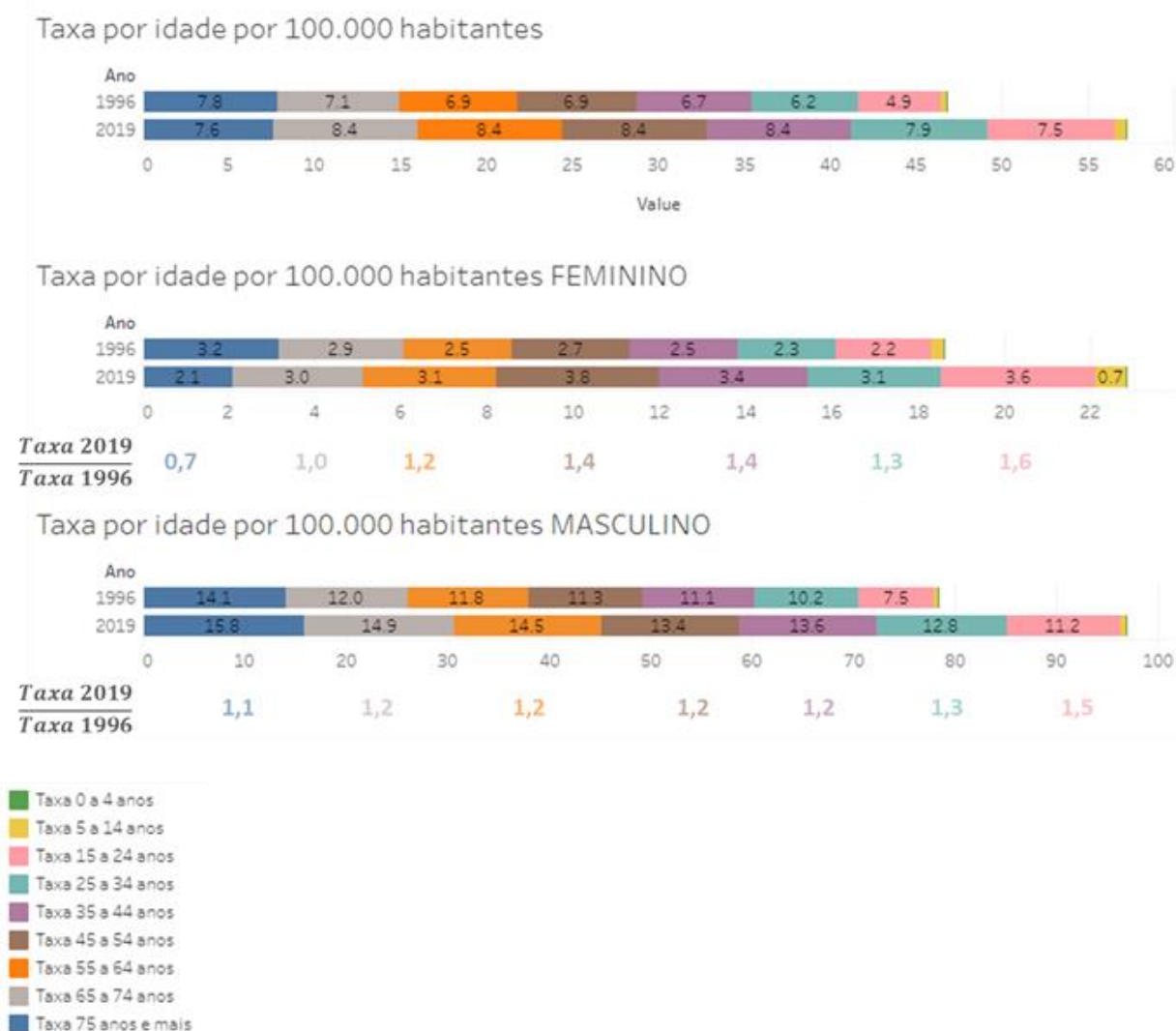


Figura 4-4 – Taxa de suicídio por 100.000 habitantes por gênero masculino/feminino e faixa de idade.

Como cidades mais populosas impactam mais a análise ponderada na taxa de suicídio do Brasil e podem causar até distorção em relação às outras cidades, deseja-se entender o suicídio nessas cidades, preliminarmente. Do total de 5.567 municípios, 25 deles (0,4% do total) representam 25% da população e 16% do total de suicídios. Pela diferença entre 25% e 16%, conclui-se, inicialmente, que o suicídio está mais pulverizado do que a concentração da população. Destaque para as cidades de Teresina (PI), Campo Grande (MS) e Porto Alegre (RS) que apresentam taxas padronizadas de suicídio maiores que 8, sendo que no Brasil a taxa resultante é 6.

Dessas 25 cidades, 48% delas apresentam aumento na taxa padronizada do suicídio por gênero masculino/feminino e idade em 2019 comparado com 1996. As cinco megacidades com os maiores aumentos da taxa de suicídio em ordem decrescente são: Campinas (SP), João Pessoa (PB), Duque de Caxias (RJ), Teresina (PI) e São Bernardo do Campo (SP).

Observa-se pela Tabela 4.1, pelo mapa de calor, que o aumento da taxa está concentrado em todas as cidades para cidadãos entre 15 a 44 anos, sendo pulverizado entre homens e mulheres. Não há nenhuma cidade anômala, em que é ela a grande responsável pelo aumento do suicídio, e as cidades seguem o mesmo padrão analisado no Brasil como um todo.

UF	Município	Taxa padronizada por sexo e idade	% acumulado - população Brasil	% acumulado - suicídio Brasil	Taxa 2019 / Taxa 1996	Aumento da taxa de 2019 comparado com os anos 1996																
						Faixa idade - MASCULINO							Faixa idade - FEMININO									
						5-14	15-24	25-34	35-44	45-54	55-64	65-74	75+	5-14	15-24	25-34	35-44	45-54	55-64	65-74	75+	
SP	SAO PAULO	1,8	6%	2%	0,3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
RJ	RIO DE JANEIRO	3,1	9%	3%	1,3	-	2	1	1	1	1	1	0	-	4	2	3	1	1	2	1	
DF	BRASILIA	6,6	10%	5%	1,0	1	1	1	1	1	1	1	-	0	3	2	1	0	1	-	0	
BA	SALVADOR	1,4	12%	5%	1,3	-	1	1	1	2	1	1	0	-	-	-	0	-	-	-	-	
CE	FORTALEZA	3,8	13%	6%	0,6	-	0	1	1	1	1	0	0	1	1	0	1	4	0	0	-	
MG	BELO HORIZONTE	6,2	14%	7%	1,2	-	1	1	1	1	1	0	-	0	4	1	1	3	4	-	1	
AM	MANAUS	4,9	15%	8%	0,8	0	1	1	0	1	1	1	0	1	1	-	-	-	-	-	-	
PR	CURITIBA	5,7	16%	9%	1,1	-	1	1	1	1	1	3	0	0	-	3	1	3	1	1	0	-
PE	RECIFE	2,8	17%	9%	0,6	-	2	0	0	1	0	2	0	-	3	0	0	-	0	1	0	
GO	GOIANIA	7,8	18%	10%	1,3	-	1	1	2	1	1	5	-	0	1	1	1	1	1	0	-	
PA	BELEM	3,7	18%	10%	0,5	1	1	0	1	0	1	1	0	1	0	1	0	-	0	0	-	
RS	PORTO ALEGRE	8,3	19%	11%	0,8	-	1	1	1	1	1	1	0	0	0	0	1	1	2	0	-	
SP	GUARULHOS	5,5	20%	12%	1,2	-	1	2	1	1	1	1	0	-	1	1	2	-	-	0	0	
SP	CAMPINAS	7,3	20%	13%	3,0	-	5	3	3	2	-	1	0	0	-	3	1	-	-	-	-	
MA	SAO LUIS	3,9	21%	13%	0,8	1	3	1	0	1	0	-	-	-	1	0	-	-	0	-	-	
RJ	SAO GONCALO	2,6	21%	13%	0,9	-	-	1	0	2	1	0	-	0	-	1	0	-	1	0	-	
AL	MACEIO	3,5	22%	13%	0,8	0	0	1	1	1	1	-	-	1	0	2	-	2	0	-	-	
RJ	DUQUE DE CAXIAS	3,4	22%	14%	1,8	0	2	1	1	1	1	-	0	-	2	4	-	-	-	-	-	
MS	CAMPO GRANDE	10,3	23%	14%	1,2	-	1	1	2	1	1	1	-	0	4	1	2	2	1	-	-	
RN	NATAL	3,8	23%	15%	0,8	-	1	2	1	1	0	0	-	0	1	0	-	0	0	0	0	
PI	TERESINA	10,4	24%	15%	1,7	-	1	4	1	1	4	1	-	-	1	-	2	1	0	-	-	
SP	SAO BERNARDO DO CAMPO	4,4	24%	16%	1,7	-	9	1	1	1	-	1	-	-	1	1	-	-	-	-	-	
RJ	NOVA IGUACU	1,1	24%	16%	0,5	-	1	1	1	0	1	1	-	-	-	0	0	-	0	-	-	
PB	JOAO PESSOA	5,2	25%	16%	2,5	-	-	1	1	-	-	-	-	-	-	-	-	-	0	-	-	
SP	SAO JOSE DOS CAMPOS	3,3	25%	16%	0,7	-	1	1	1	0	1	1	0	-	0	1	-	0	-	-	-	

Tabela 4.1 – Taxa padronizada do suicídio e comparação do aumento das taxas de 2019 e 1996 nas 25 primeiras cidades mais populosas do Brasil.

Inicialmente, os municípios destacados não possuem um padrão comum entre si. Isso pode levar à conclusão inicial de que cada região brasileira deve ser estudada

separadamente, com o apoio deste material. Além disso, observa-se inicialmente que não são todas as cidades que apresentam o mesmo movimento de crescimento da taxa; então, não se pode generalizar e cada caso deve ser avaliado.

Após a análise em grandes números do suicídio no Brasil e nas grandes cidades, deseja-se entender melhor como o suicídio se apresenta em cada município brasileiro.

Para se analisar individualmente cada município, é indicado inicialmente fazer o tratamento dos casos de municípios que são intermitentes e que apresentam uma pequena população, pois como não há a suavização espacial, eles podem apresentar anomalias, conforme citado no Item “3.3 – VALIDAÇÃO DOS DADOS E ANÁLISE DESCRITIVA”. Essa anomalia pode ser chamada de “efeito vagalume”.

Buscou-se primeiramente agrupar as séries temporais da taxa de suicídio de cada cidade para se retirar os agrupamentos de cidade que apresentassem o efeito vagalume citado acima. Quando se analisa o gráfico da somatória da distância quadrada de cada ponto até seu centróide para cada número de agrupamentos realizado, não é encontrado um ponto de inflexão (*elbow*), conforme ilustra a Figura 4-5. Portanto não há um número ideal de *clusters* para a medida, e esse não é o melhor método para se poder tirar os *outliers*.

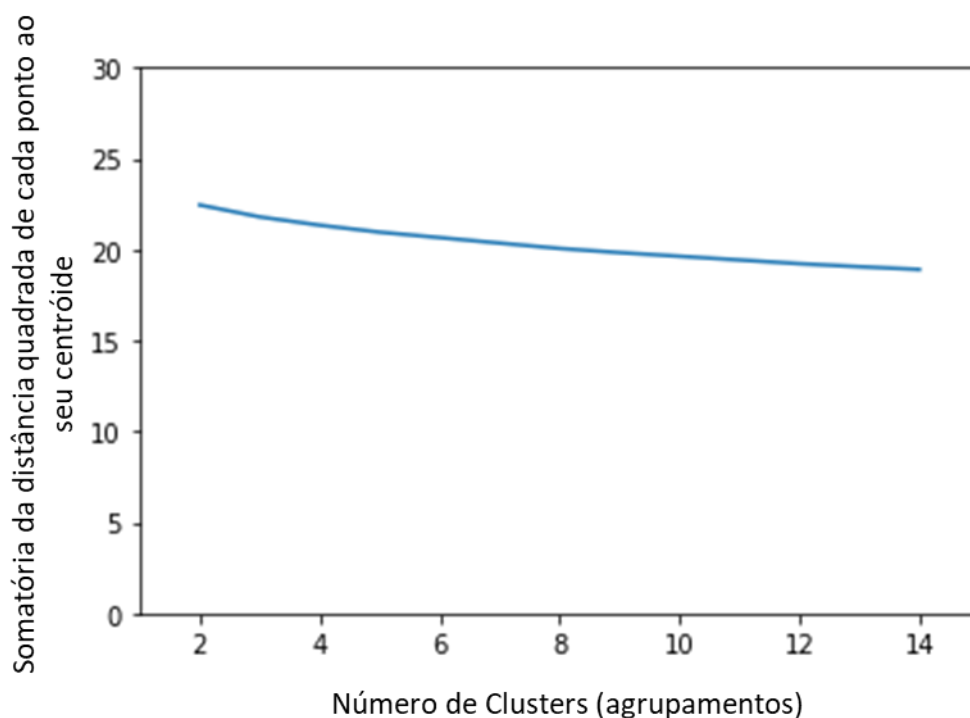


Figura 4-5 – Somatória da distância quadrada de cada ponto ao seu centróide a cada número de clusters determinado. Não se enxerga nessa figura o *Elbow*.

Foi feito o teste com até 30 *clusters* e não se observou ponto de inflexão na curva do *elbow*. Abaixo está a imagem apenas ilustrativa de 14 *clusters* das séries temporais do suicídio das cidades (Figura 4-6). Cada cluster é um gráfico, cada linha preta é a série temporal de uma cidade e a linha vermelha é a série temporal do centróide do cluster. Observa-se que há agrupamentos com tendências de crescimento (linha vermelha subindo ao longo dos anos), outros com tendência de decréscimo (linha vermelha descendo ao longo do eixo x) e outros estáveis. Observa-se também que há bastante ruído (diversas linhas pretas para cima e para baixo da linha vermelha) e o *elbow* se justifica. Essa resposta não está confiável e não é possível se retirar as cidades “vagalume” utilizando-se a técnica de agrupamento.

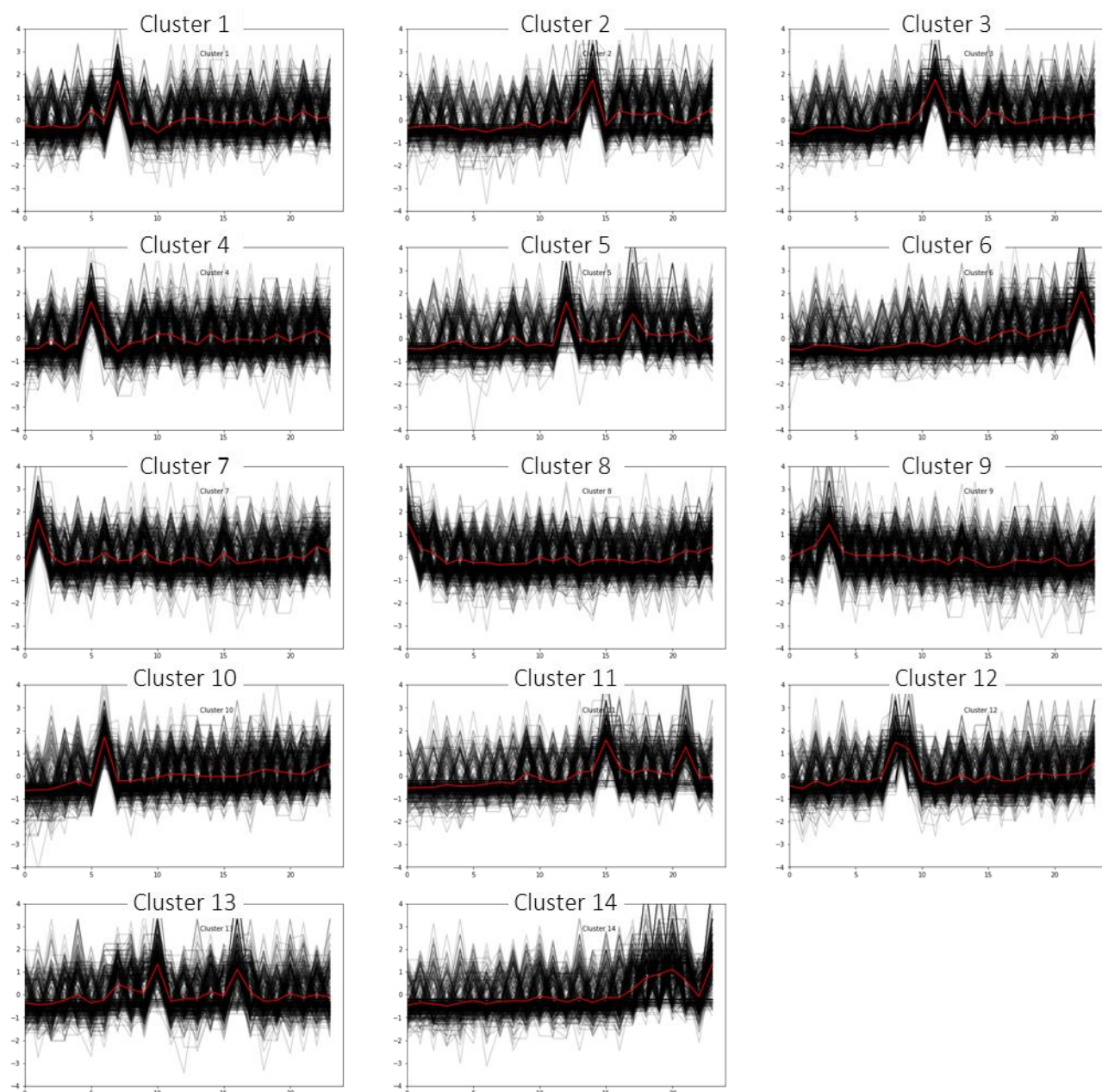


Figura 4-6 – Ilustração dos agrupamentos das séries temporais obtidos da taxa de suicídio entre os municípios. Cada gráfico é um agrupamento, o eixo x está o tempo (anos) e o eixo y é a taxa de suicídio normalizada. As linhas pretas são as séries temporais de cada município e a linha vermelha é a série temporal do centróide do agrupamento.

Caso houvesse um número de agrupamentos ideal para tratar o caso, provavelmente seria possível visualizar baixo ruído nos gráficos apresentados anteriormente, ou seja, as linhas pretas estariam próximas das linhas vermelhas. Além

disso, seria possível visualizar todas essas linhas crescendo ou diminuindo juntas ao longo do eixo x (anos).

Como não foi obtido sucesso no agrupamento, conforme descrito no Item “3.3 – VALIDAÇÃO DOS DADOS E ANÁLISE DESCRITIVA”, adota-se a taxa de suicídio igual a 13 para casos maiores que 13 para ser possível analisar os mapas e não haver *outliers* atrapalhando a visualização.

Os mapas da Figura 4-7 até a Figura 4-13 auxiliam inicialmente no entendimento do suicídio em todo o território e em uma análise inicial de cada cidade individualmente.

No mapa da Figura 4-7, observa-se que há municípios com altas taxas de suicídio em todos os estados do Brasil. Visualmente, notam-se alguns possíveis agrupamentos espaciais que são confirmados nas análises dos próximos itens.

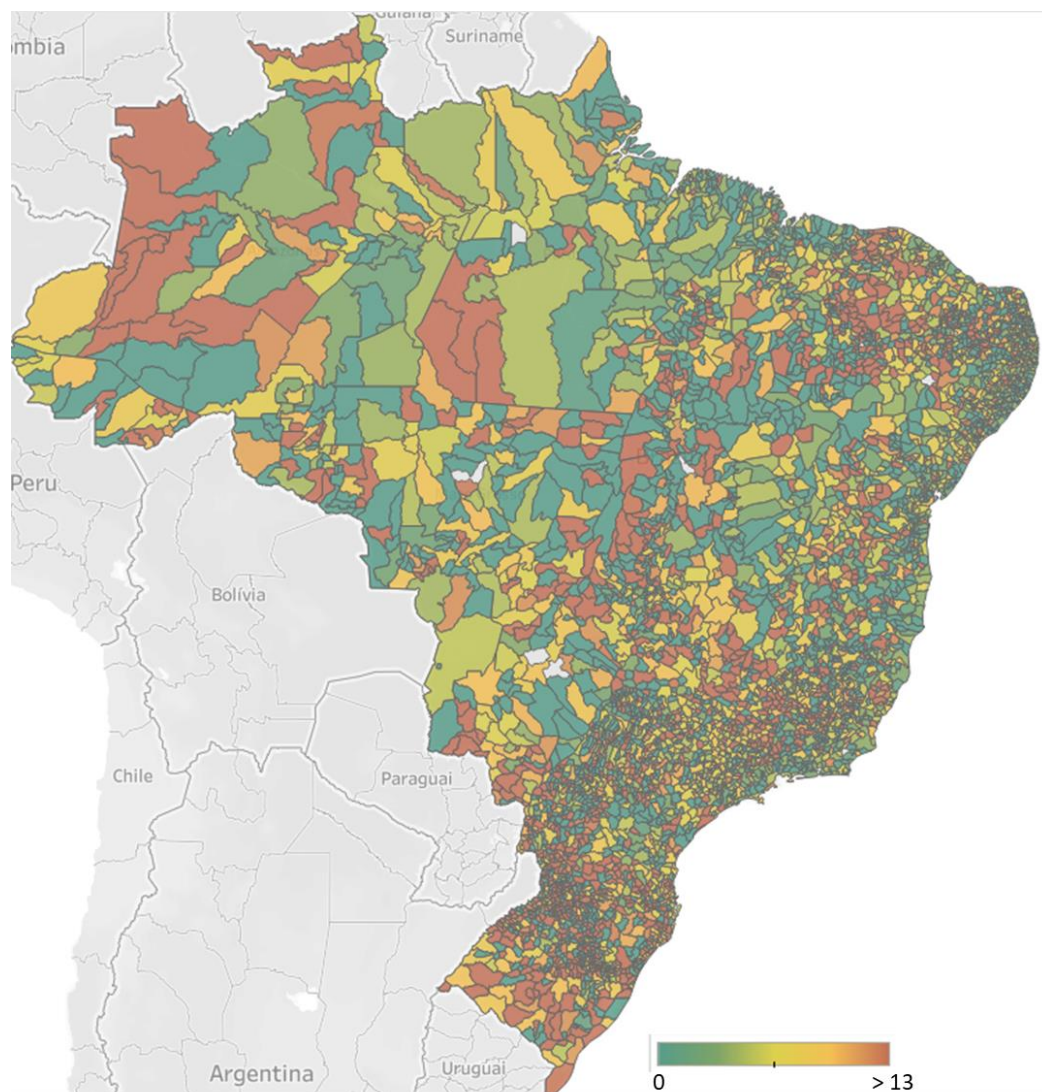


Figura 4-7 – Mapa coroplético da taxa de suicídio padronizada por gênero masculino/feminino e idade no Brasil em 2019 por município.

A Figura 4-8 compara as taxas de suicídio entre 1996, 2010 e 2019. Observa-se visualmente o crescimento da taxa de suicídio ao longo dos anos no Brasil em todas as regiões, o mapa vai se tornando mais alaranjado ao longo do tempo. É necessário aumento de escala para analisar a região Sudeste e para aprofundamento do que acontece em cada cidade do Brasil.

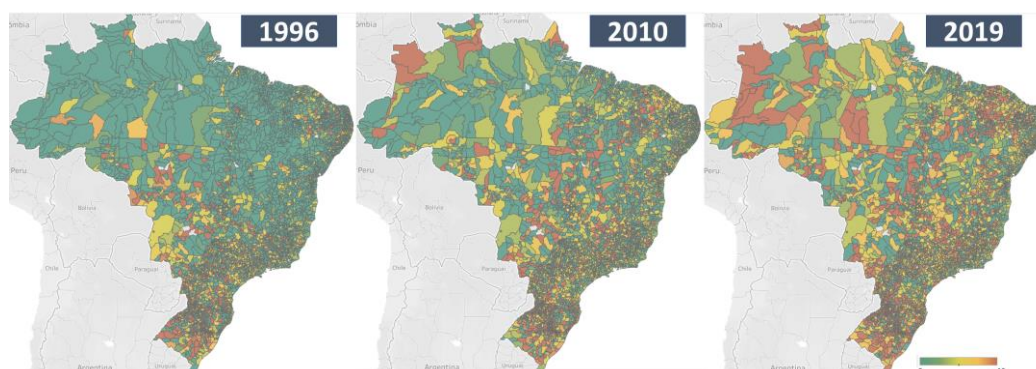


Figura 4-8 – Mapa Coroplético da taxa padronizada de suicídio por gênero masculino/feminino e idade nos municípios brasileiros em 1996, 2010 e 2019.

Muitas das cidades que apresentam altas taxas podem ser as cidades com pequenas populações, que podem causar distorções nos números por falta da suavização espacial. Nos aprofundamentos sobre quais cidades há aumento ou diminuição da taxa, são consideradas apenas cidades com suicídio em ao menos 75% dos anos, conforme descrito em “3.3 – VALIDAÇÃO DOS DADOS E ANÁLISE DESCRITIVA”. Ou seja, cidades que registraram suicídio em 18 anos, do total de 24 anos analisados. Assim, 1.214 municípios são considerados nas próximas análises, que são 21% do total de 5.659 municípios.

Ao se aumentar as escalas das regiões, comparando 2010 com 2019, observa-se visualmente em que cidades ocorreu o aumento da taxa:

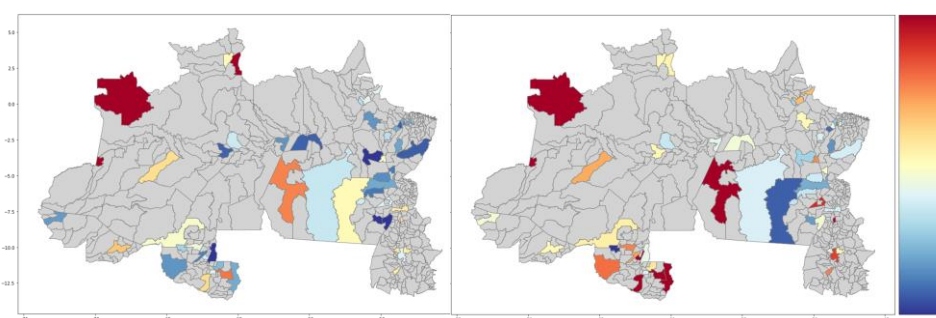


Figura 4-9 - Região Norte: comparação de 2010 esquerda e 2019 direita. Quanto mais vermelha a cor, maior a taxa de suicídio, quanto mais azul, menor a taxa.

Na região Norte, observa-se aumento em uma cidade do Pará (Itaituba), mas ao mesmo tempo diminuição em outra (São Félix do Xingu). No Amazonas, ponto de

atenção para o município de São Gabriel da Cachoeira. Em Rondônia, destaque para os municípios que são fronteiras com outros estados ou países.

No Nordeste, destaque para o estado do Ceará e logo abaixo dele. Nos próximos itens, mais detalhes são abertos por cidades e na ferramenta em Dada (2022) é possível analisar cada cidade individualmente.

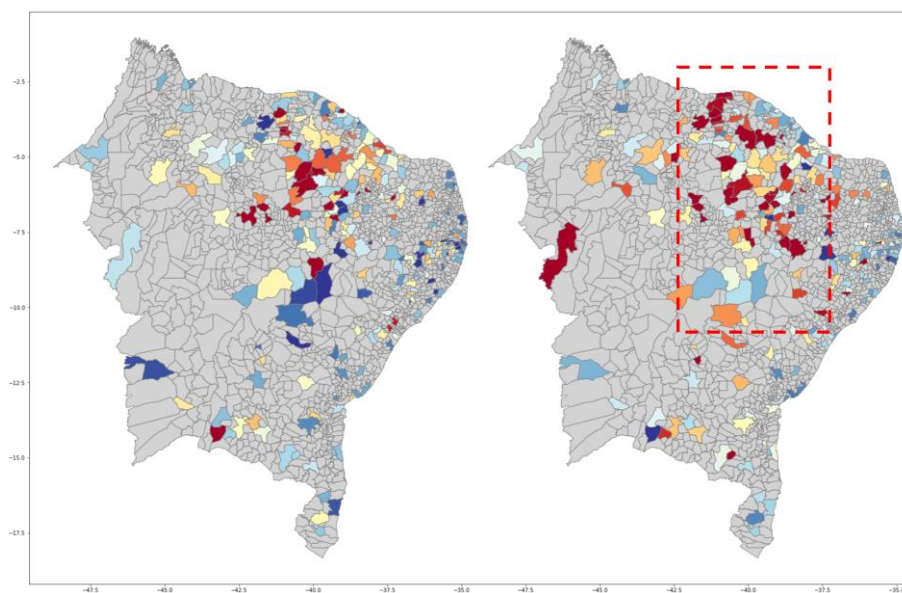


Figura 4-10 - Taxa de suicídio nos municípios da região Nordeste em 2010 (esquerda) e 2019 (direita). Quanto mais vermelha a cor, maior a taxa de suicídio, quanto mais azul, menor a taxa.

No Sudeste destaque para o Sudoeste de Minas e o Centro de São Paulo.

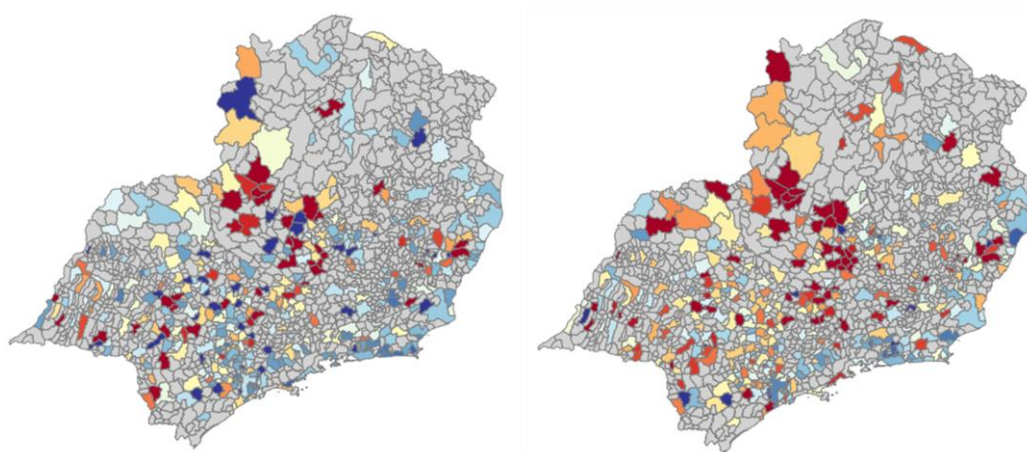


Figura 4-11 – Taxa de suicídio nos municípios da região Sudeste em 2010 (esquerda) e 2019 (direita). Quanto mais vermelha a cor, maior a taxa de suicídio, quanto mais azul, menor a taxa.

No Sul, destaque para o extremo Sul do país.

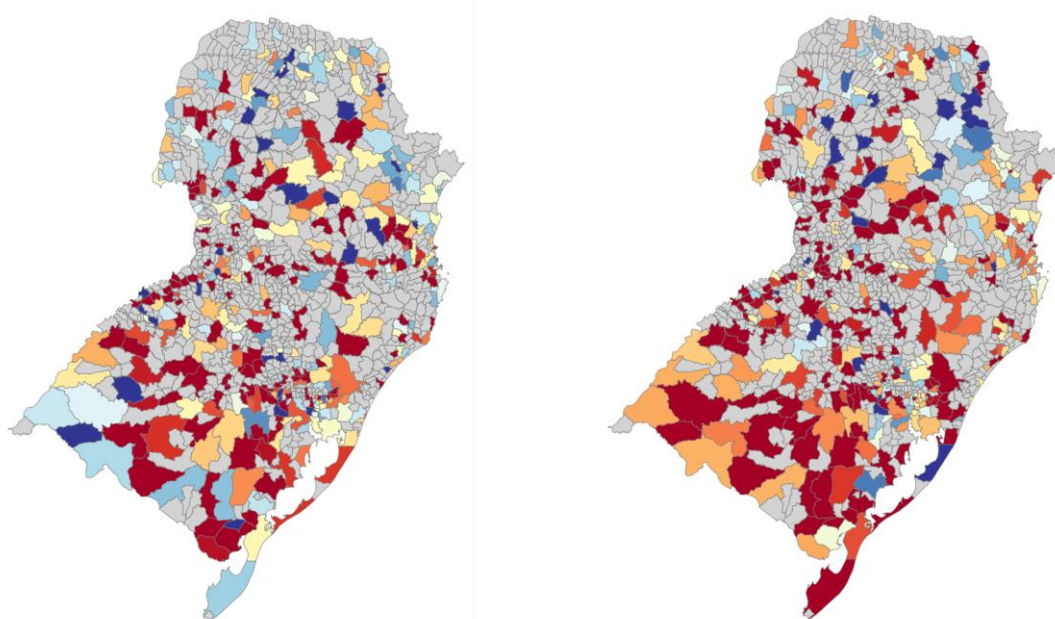


Figura 4-12 – Taxa de suicídio nos municípios da região Sul em 2010 (esquerda) e 2019 (direita). Quanto mais vermelha a cor, maior a taxa de suicídio, quanto mais azul, menor a taxa.

No Centro-Oeste, destaque para o norte do Mato Grosso, Sul de Goiás e Sul de Mato Grosso do Sul.

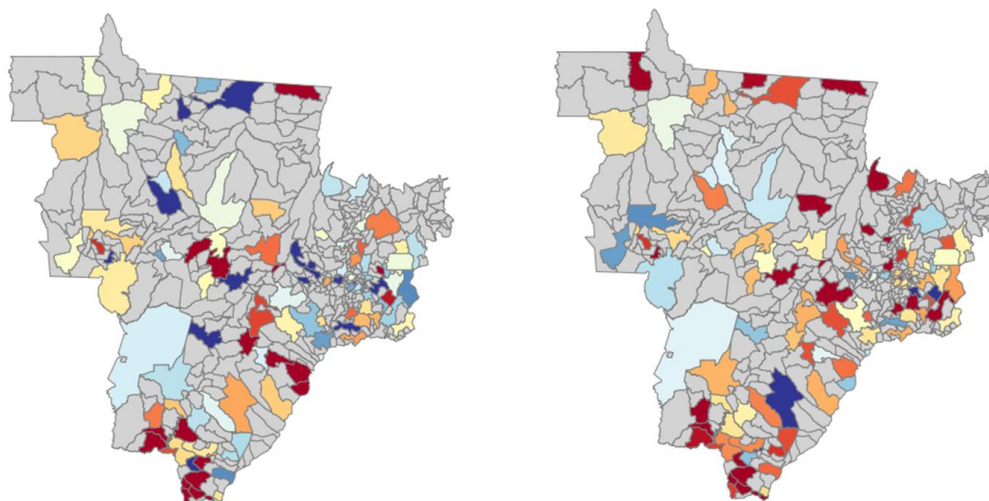


Figura 4-13 – Taxa de suicídio nos municípios da região Centro-Oeste em 2010 (esquerda) e 2019 (direita). Quanto mais vermelha a cor, maior a taxa de suicídio, quanto mais azul, menor a taxa.

No geral, em uma análise visual, observa-se uma alta aglomeração de suicídios ou aumento de suicídios em:

- Região Norte:
 - Amapá;
 - Rondônia.
- Região Nordeste:
 - Ceará e Norte do sertão Nordestino;
 - Destaque para a cidade de Balsas no Maranhão, em que aumentou consideravelmente.
- Região Sudeste:
 - Sudoeste de Minas Gerais;
 - Centro de São Paulo.
- Região Sul
 - Todos os estados;
 - Rio Grande de Sul, próximo à fronteira, destaque para o crescimento de suicídio.

- Região Centro-Oeste:
 - Norte do Mato Grosso;
 - Sul Mato Grosso do Sul;
 - Sul de Goiás.

Nos painéis em Dada (2022) é possível analisar o comportamento do suicídio em cada ano de cada cidade brasileira no mapa ou em gráficos para se entender o comportamento por gênero masculino/feminino e faixa de idade ao longo do tempo.

Após a análise descritiva e conclusões preliminares, é apresentada a análise de quais são os agrupamentos espaciais significativos de alto e baixo risco no Brasil.

4.2 ANÁLISE ESPACIAL

Após a análise espacial, verificou-se que há cinco agrupamentos com alto risco de suicídio e seis agrupamentos com baixo risco, conforme Figura 4-9.

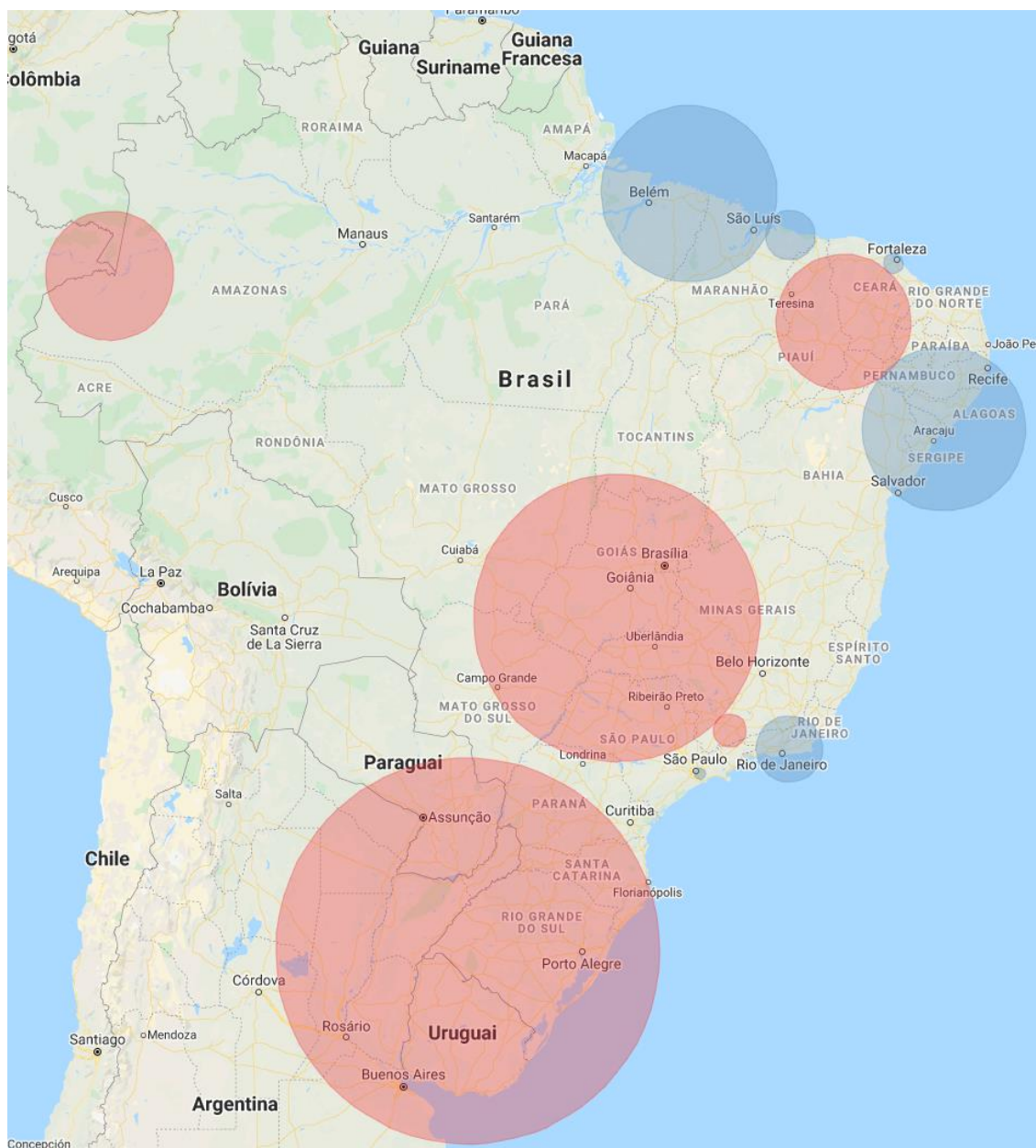


Figura 4-14 – Agrupamentos de alto risco, em vermelho, e de baixo risco, em azul.

A seguir apresentamos com mais detalhes os agrupamentos de alto risco englobam as regiões:

- Norte, com um agrupamento na divisa do Brasil com Peru e Colômbia, com a inclusão das cidades de São Paulo de Olivença, Amaturá e Benjamin Constant.

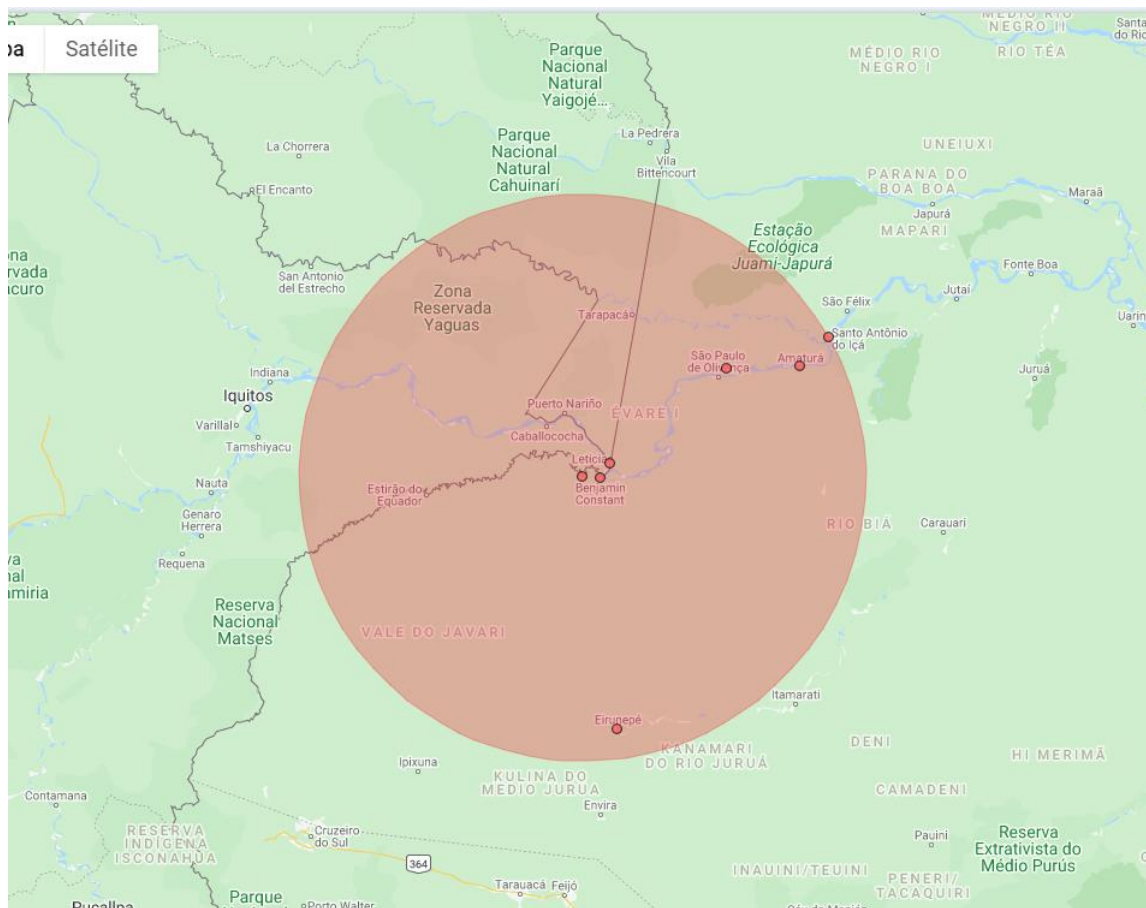


Figura 4-15 – Agrupamento de risco elevado na região Norte na divisa do Brasil com Peru e Colômbia.

- Nordeste, com um agrupamento no interior do Ceará com divisas do Rio Grande do Norte, Paraíba e Pernambuco e região centro-leste do Piauí. (parte da região do sertão brasileiro).

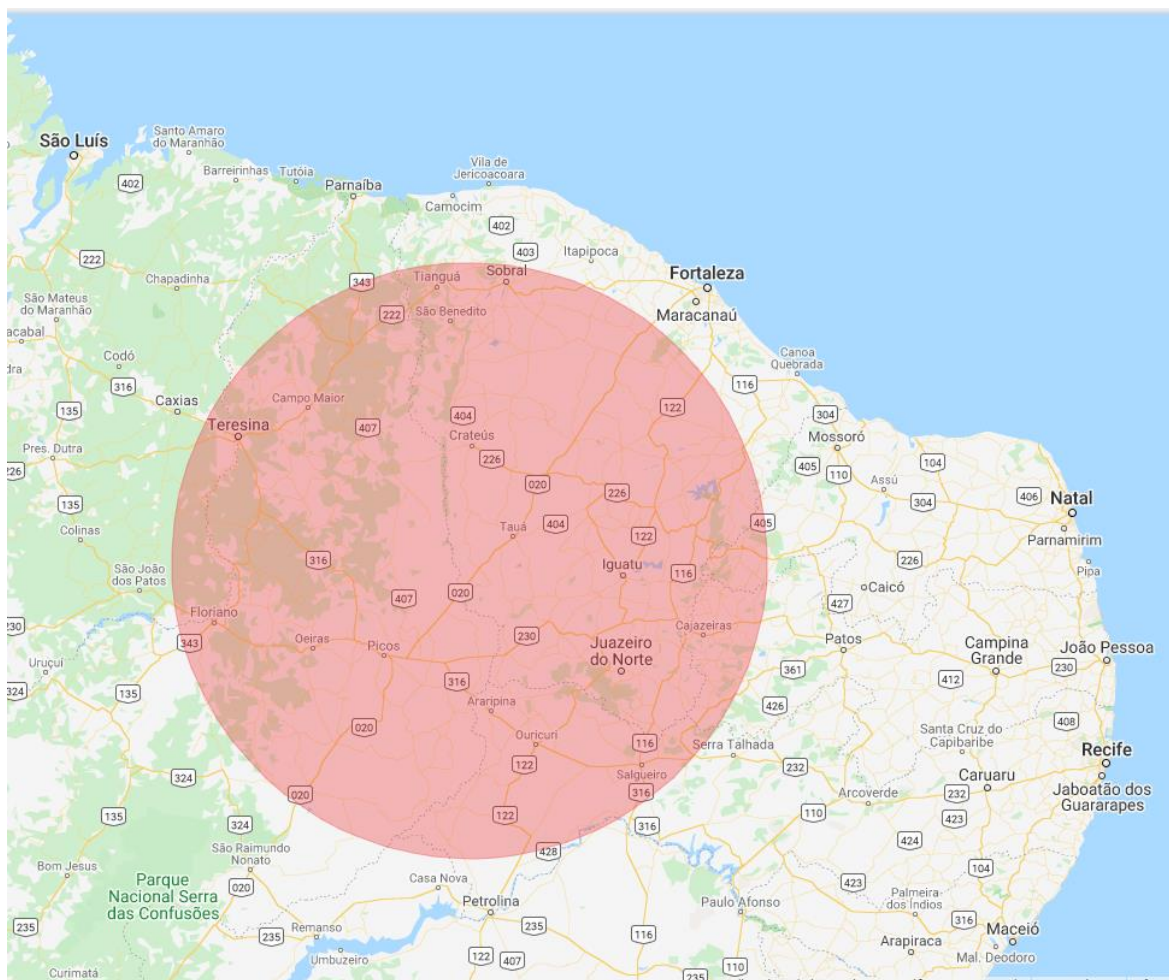


Figura 4-16 – Agrupamento de risco elevado na região Nordeste.

- No Centro-Oeste e Sudeste, o agrupamento engloba: (1) um agrupamento no interior de São Paulo, normalmente cidades com mais de 400 km de distância do litoral; (2) nordeste do Mato Grosso do Sul, com divisa no Mato Grosso, Goiás e São Paulo; (3) estado todo de Goiás; Mato Grosso divisa com Goiás e (4) oeste de Minas Gerais após Belo Horizonte.

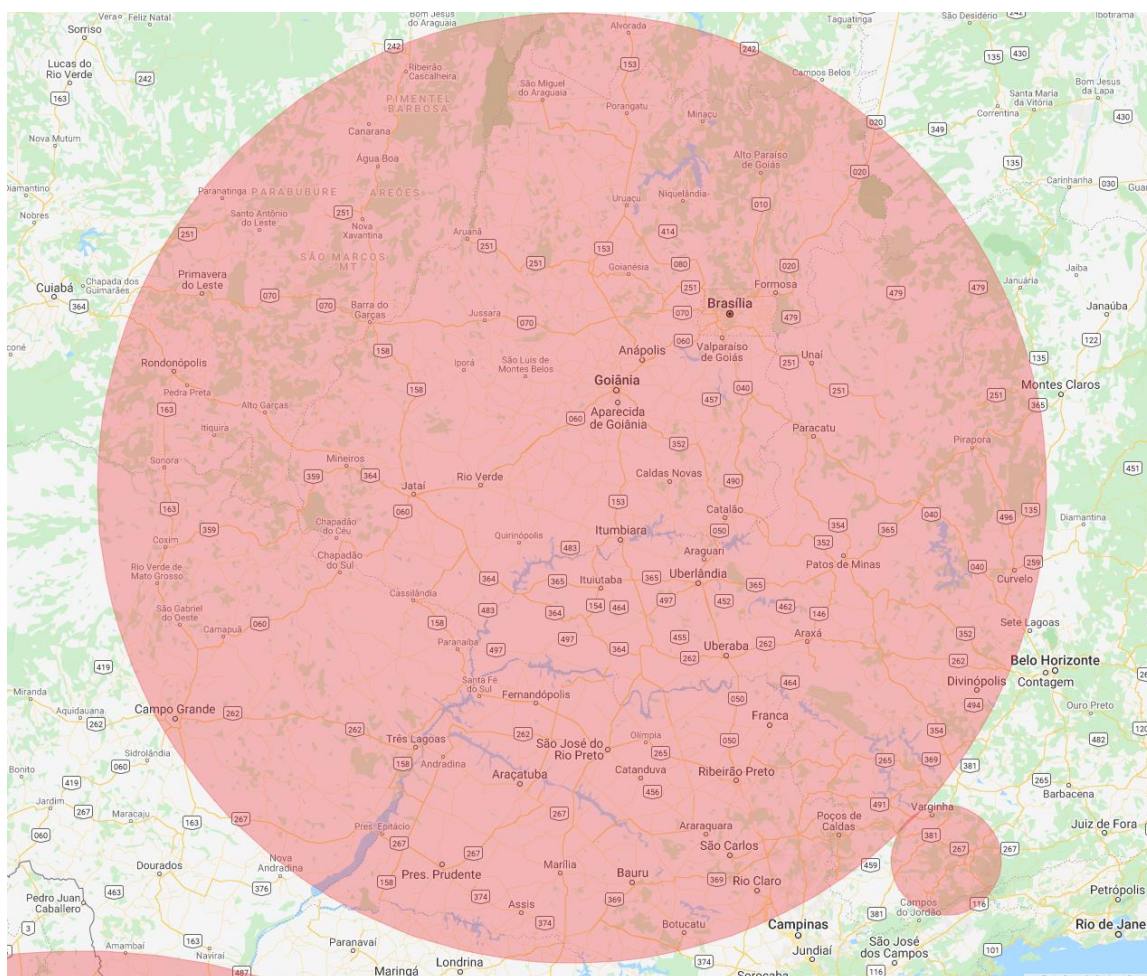


Figura 4-17 – Agrupamento de alto risco na região Centro-Oeste e Sudeste.

- Parte do Sul de Minas, com um agrupamento na região do circuito das águas.

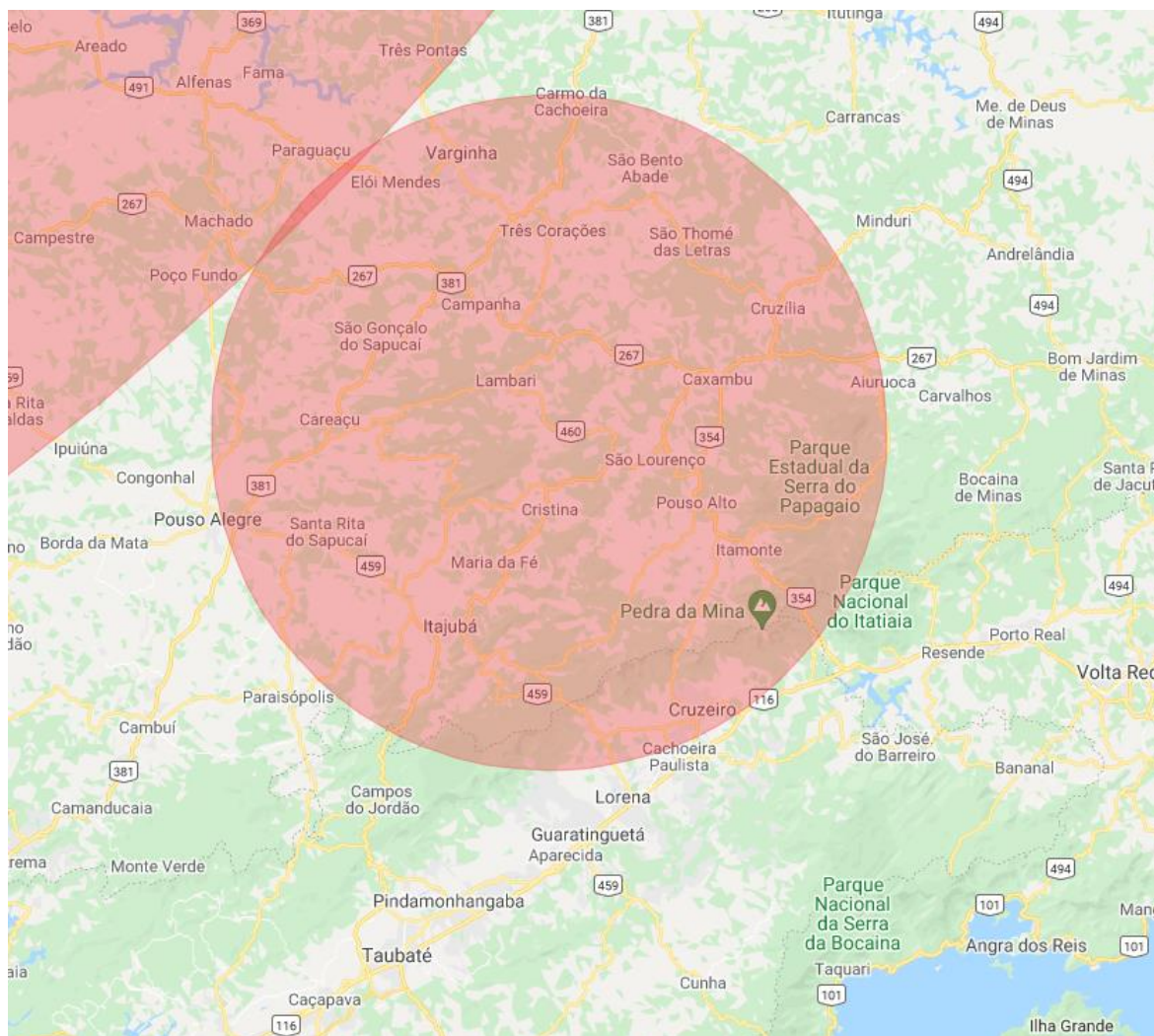


Figura 4-18 – Agrupamento de alto risco na região sul de Minas Gerais.

- Sul, com um agrupamento nos estados do Rio Grande do Sul e Santa Catarina inteiros, sul do Paraná e divisa do Paraná com o Paraguai. Embora o círculo ultrapasse as fronteiras do Brasil, a conclusão não se aplica aos países vizinhos.

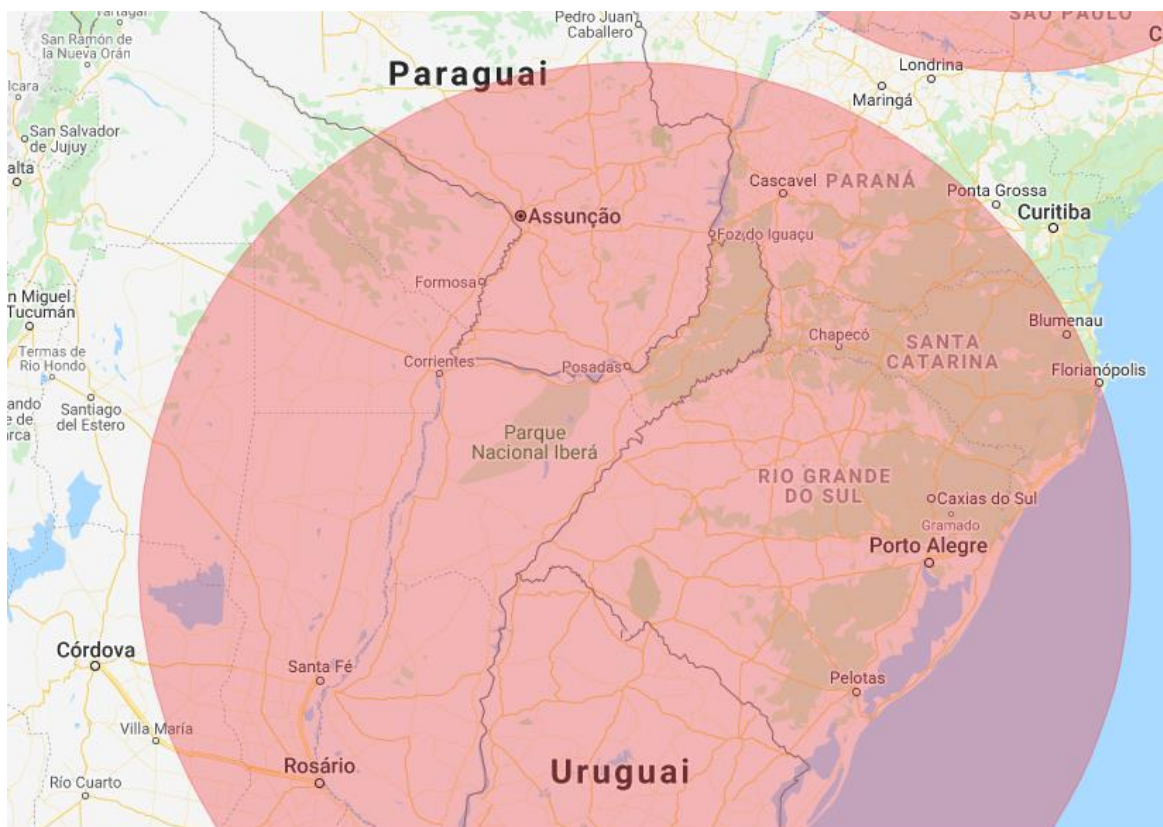


Figura 4-19 – Agrupamento de alto risco na região sul do Brasil., a conclusão não se aplica aos países vizinhos.

Os agrupamentos de baixo risco englobam principalmente as regiões com capitais litorâneas, são eles:

- Um agrupamento no Norte do Pará e Maranhão, envolvendo as capitais e arredores.

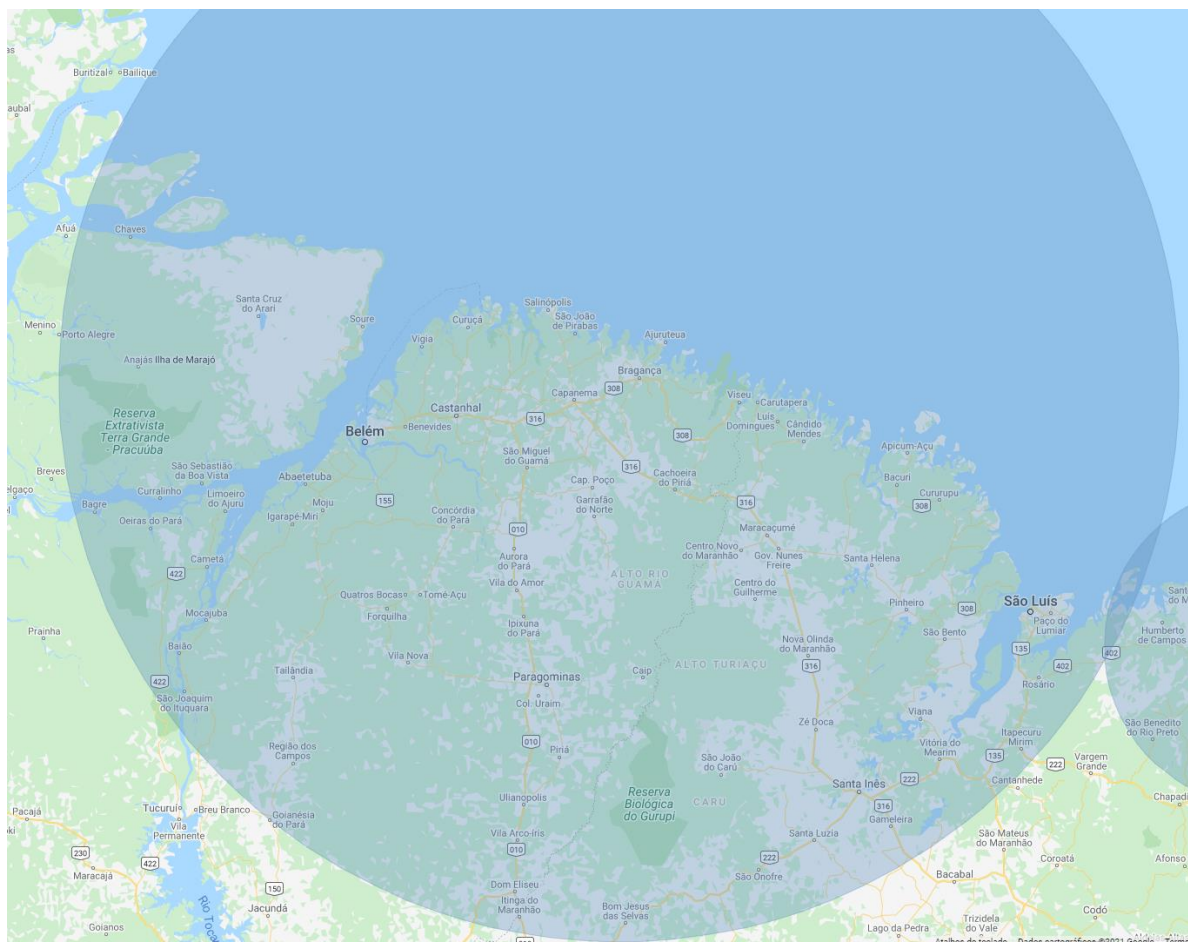


Figura 4-20 – Agrupamento de baixo risco no Pará e Maranhão.

- Um agrupamento na região litorânea nordeste do Maranhão, ao lado da cidade de São Luís.

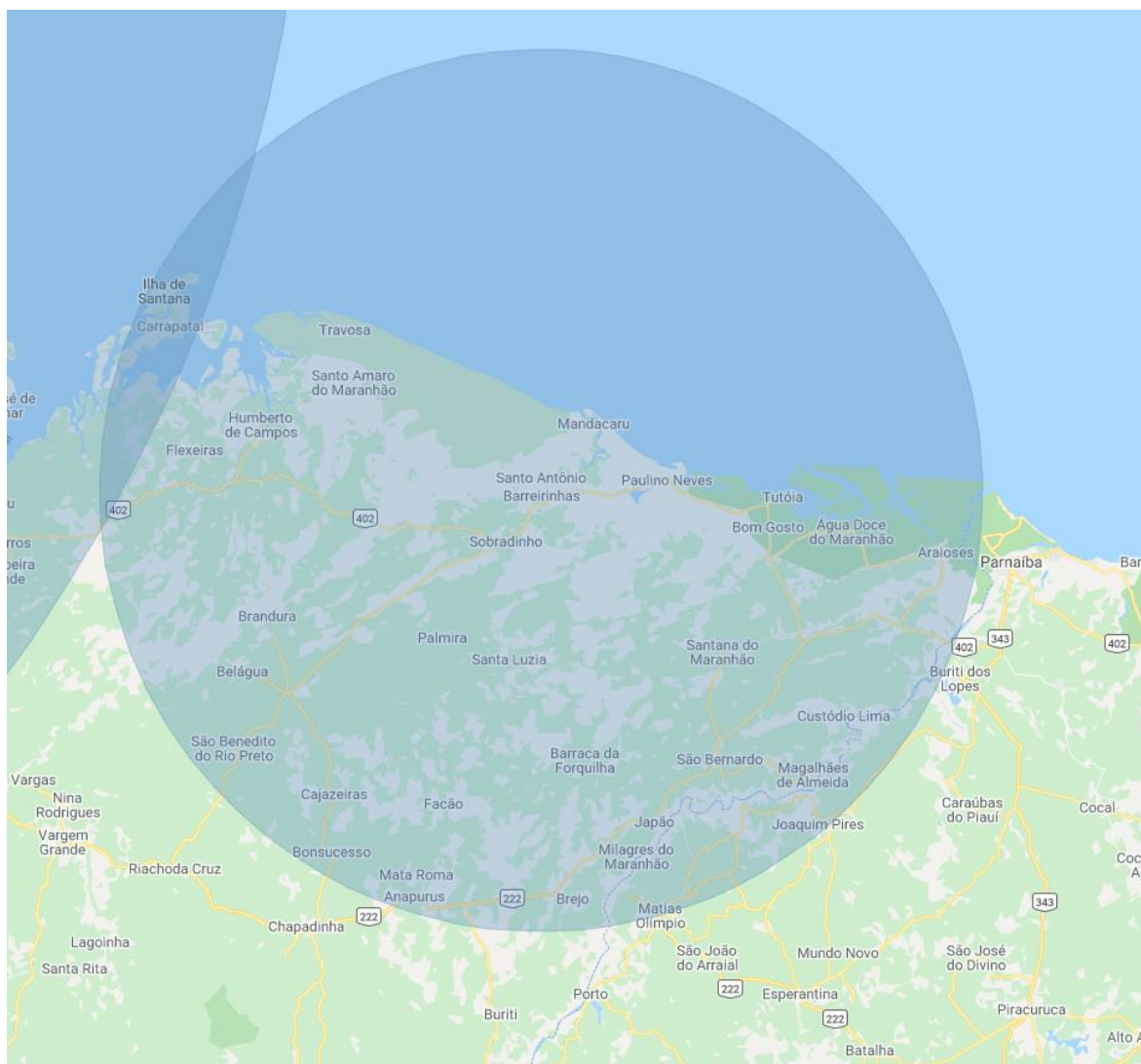


Figura 4-21 – Agrupamento de baixo risco no Maranhão

- Um agrupamento de baixo risco em Fortaleza e arredores.

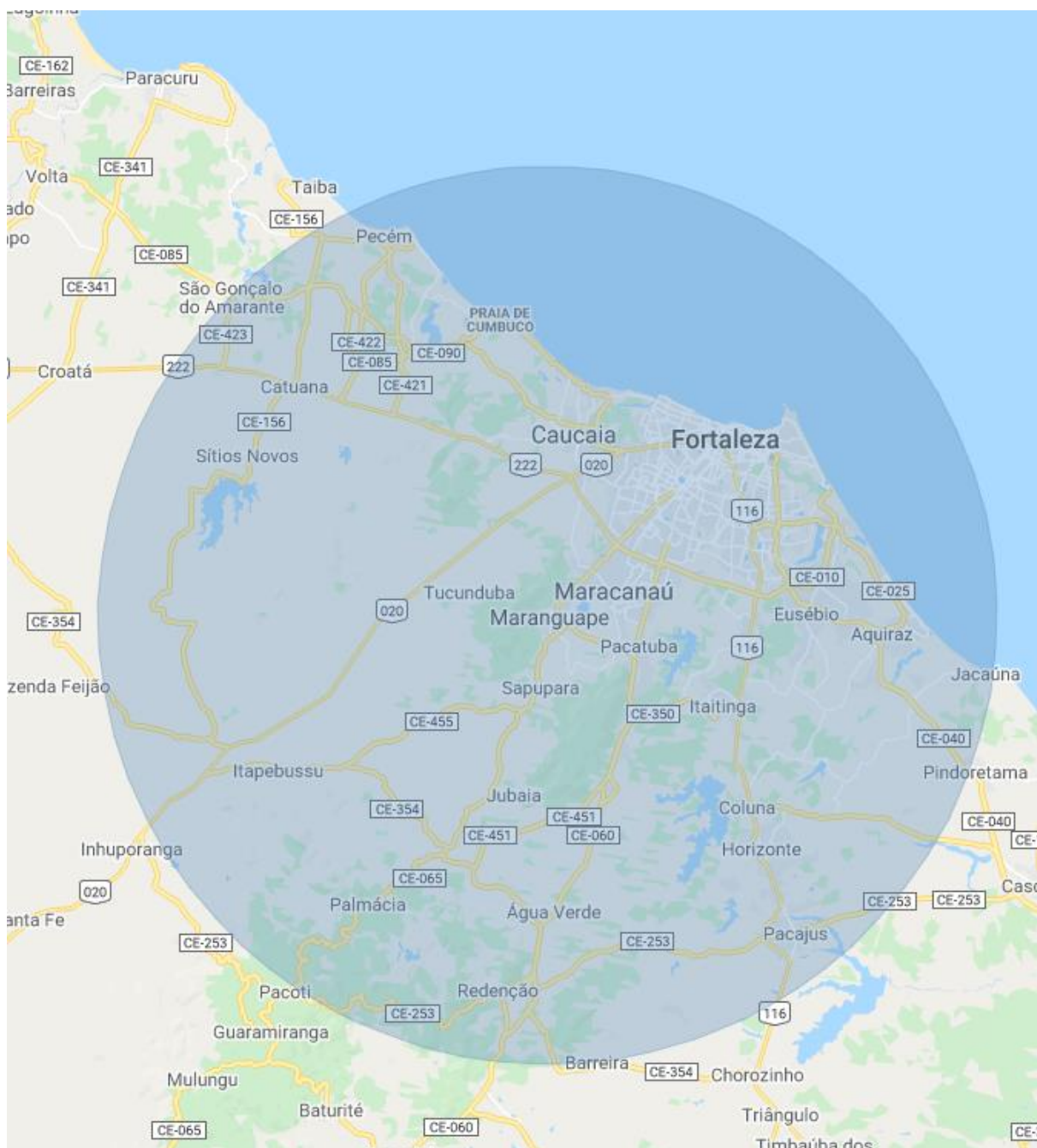


Figura 4-22 – Agrupamento de baixo risco em Fortaleza e arredores.

- Um agrupamento de baixo risco que engloba o nordeste da Bahia, a partir de Salvador, os estados de Sergipe e Alagoas inteiros e a região central e leste de Pernambuco.

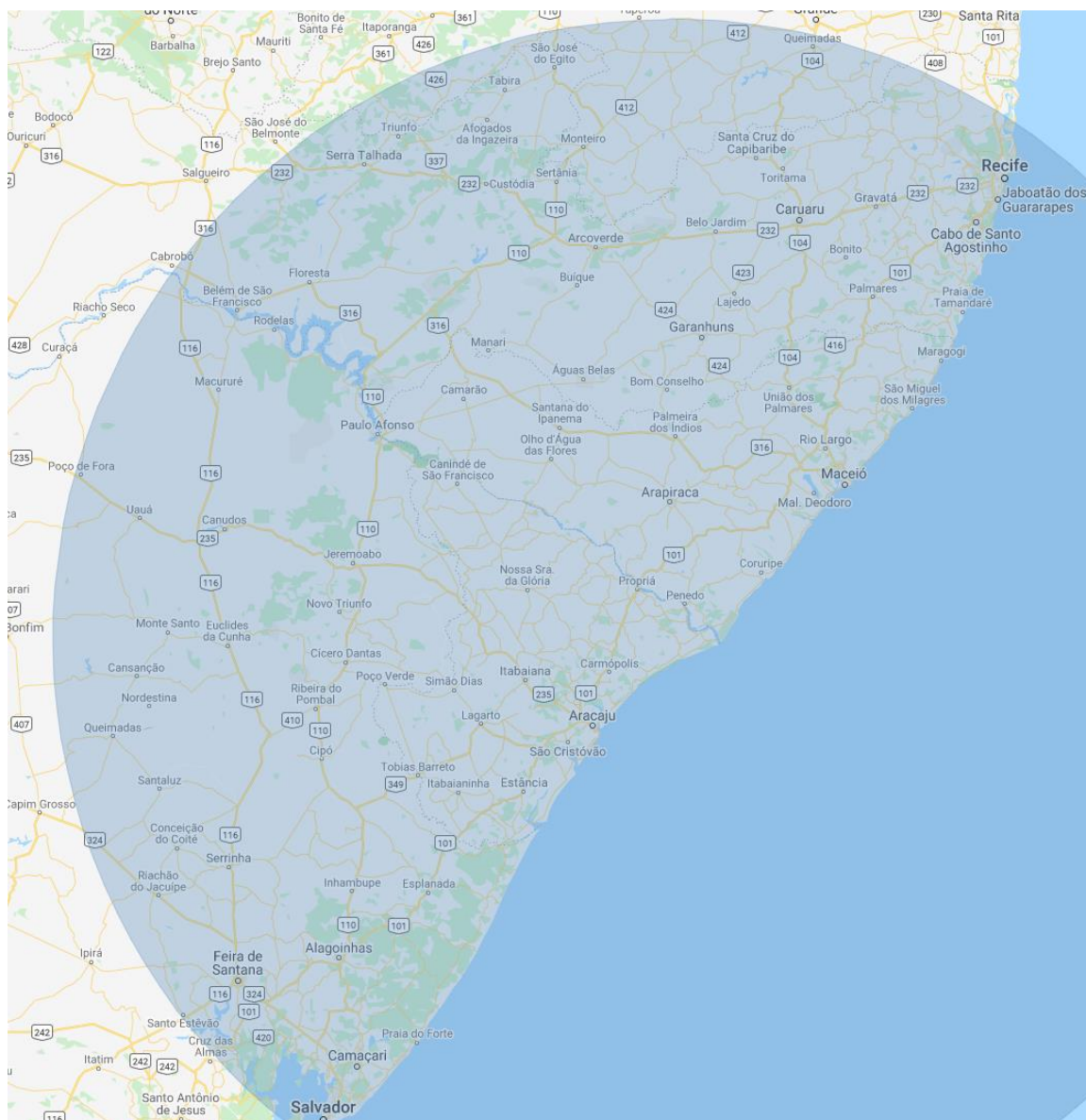


Figura 4-23 – Agrupamento de baixo risco no litoral do nordeste brasileiro e arredores.

- Um agrupamento que engloba quase todo o estado do Rio de Janeiro, com exceção do norte, e a região de Minas Gerais fronteira com o Rio de Janeiro, englobando Juiz de Fora.

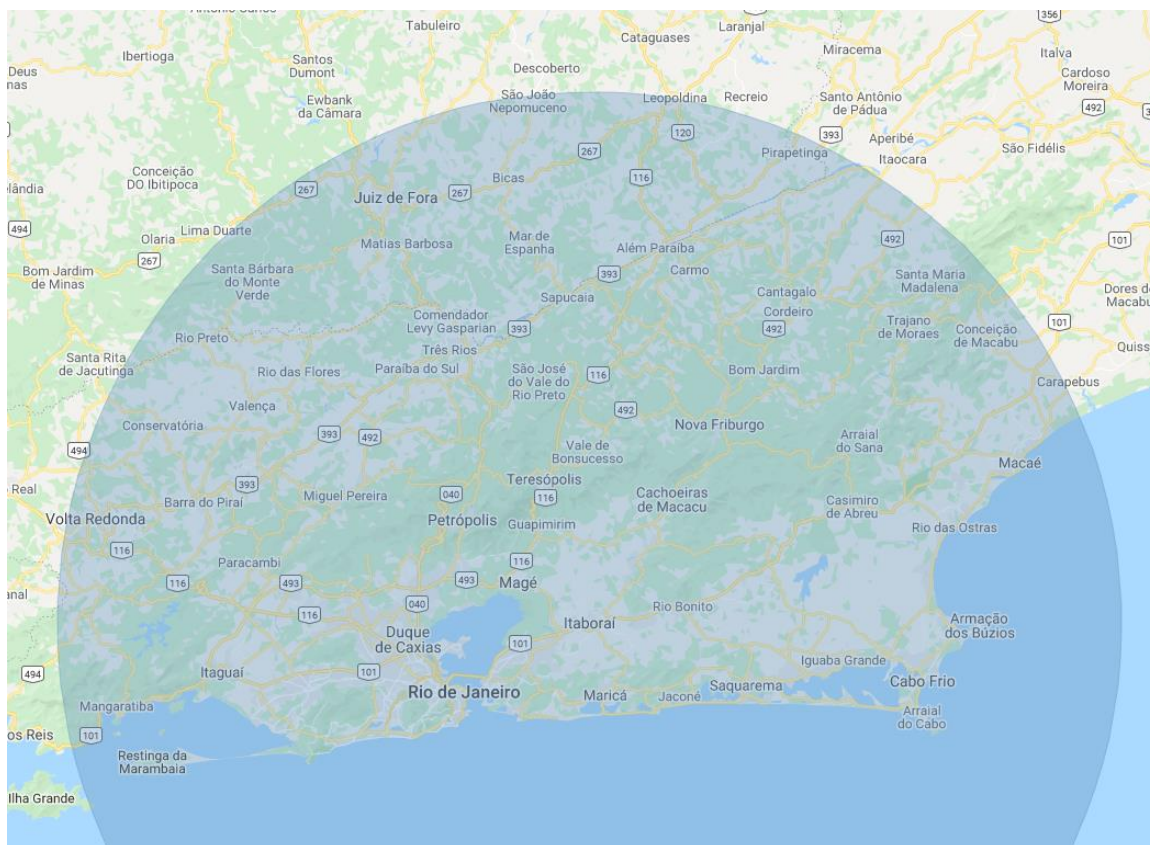


Figura 4-24 – Agrupamento de baixo risco no Rio de Janeiro.

- Um agrupamento de baixo risco que engloba a área metropolitana de São Paulo, na região sul e leste.

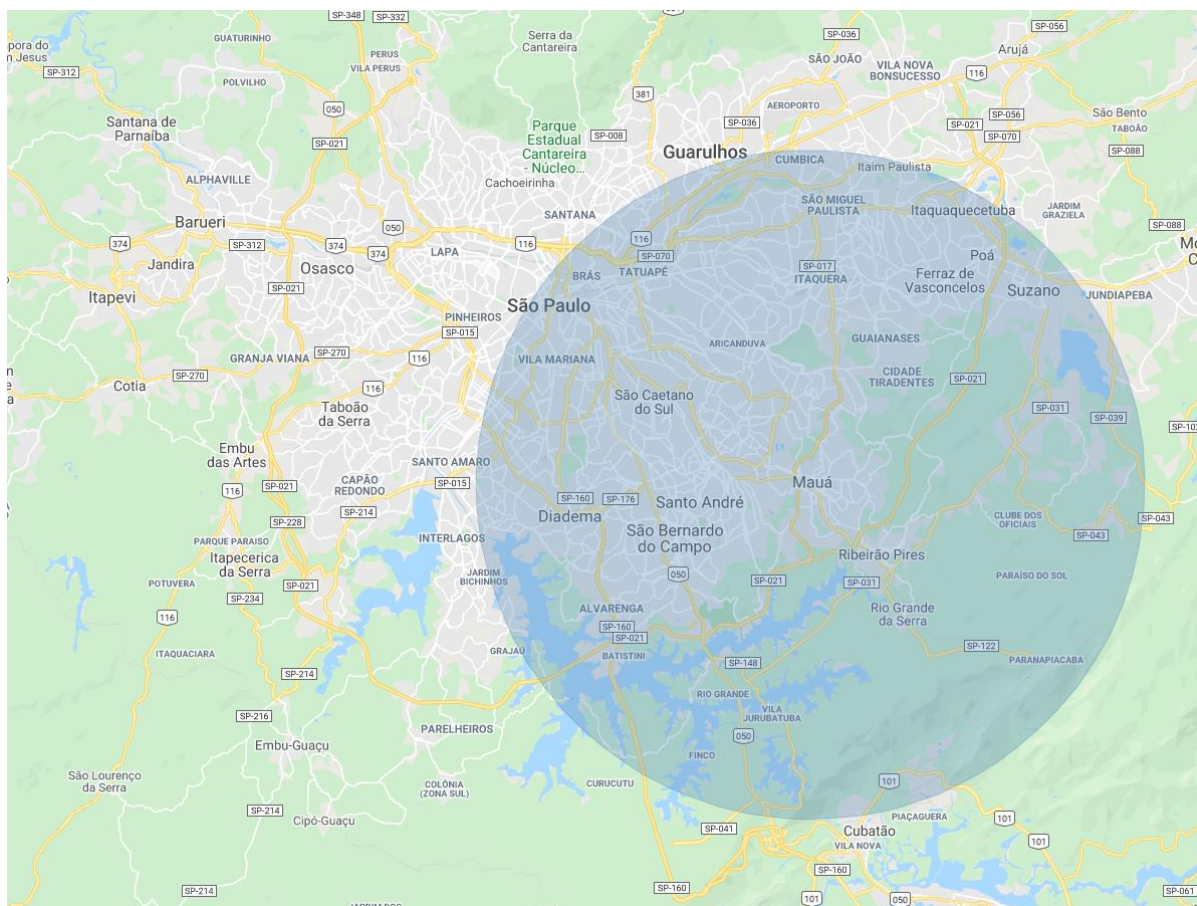


Figura 4-25 – Agrupamento de baixo risco na região metropolitana de São Paulo.

Os agrupamentos apresentados são referentes ao ano de 2019. Considerando uma mescla entre a análise espacial e temporal, deseja-se entender se esses agrupamentos se alteraram na série histórica ou não. Para isso, analisa-se os agrupamentos espaço-temporais propostos por Kulldorff (2005). Quando se analisa espaço-temporalmente, observaram-se doze agrupamentos, sendo que metade é de alto risco e metade de baixo risco. A

Figura 4-26 ilustra esses agrupamentos.

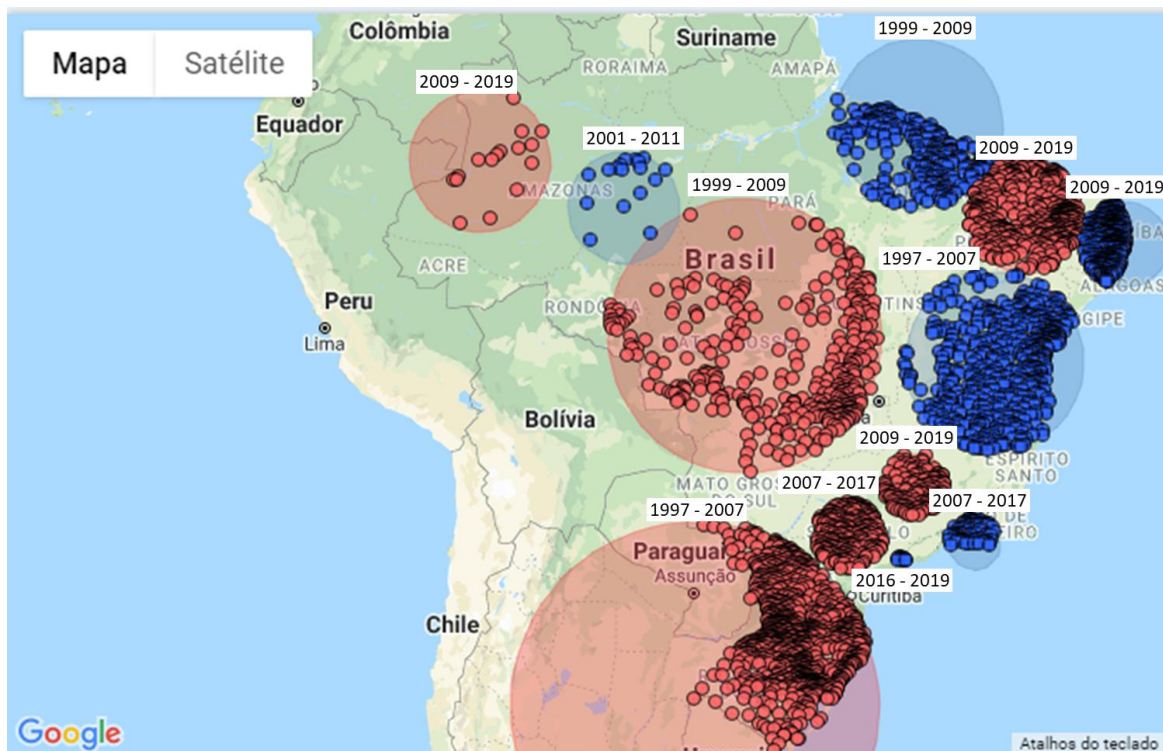


Figura 4-26 – Agrupamentos espaço-temporais de alto risco (vermelho) e baixo risco (azul) no Brasil de 1997 a 2019. Cada círculo pequeno representa um município.

Nos agrupamentos de alto risco, o agrupamento espaço-temporal apresenta tendências similares aos agrupamentos espaciais de 2019, porém com algumas discrepâncias, que são comentadas a seguir.

Observa-se que o agrupamento da região Norte do Brasil engloba na análise espaço temporal mais cidades interioranas, além das cidades que são fronteiras do Peru e Colômbia. Isso demonstra que a taxa pode estar regredindo nas regiões interioranas.

Na região Nordeste, além do agrupamento existente em 2019, observa-se que historicamente a região litorânea do Ceará também é considerada de alto risco e atualmente ela não está mais incluída no *cluster*, podendo indicar que a taxa pode estar regredindo na localidade.

No Centro-Oeste, havia um agrupamento de alto risco de 1999 a 2009 que englobava quase toda a totalidade da região e o sul do Pará. Há também um agrupamento no Sudoeste de Minas Gerais entre as cidades (exclusive) de Belo Horizonte, Uberlândia, Uberaba e Ribeirão Preto e outro grupo na região Oeste do

estado de São Paulo entre as cidades (exclusive) de Presidente Prudente, São José do Rio Preto, Ribeirão Preto, Sorocaba e Londrina Na análise puramente espacial mais atual, esses três agrupamentos regrediram e atualmente, esse grupo está mais concentrado nas fronteiras entre o Tocantins, Mato Grosso, Mato Grosso do Sul, São Paulo e Minas Gerais.

Na região Sul, historicamente o agrupamento é maior do que o agrupamento espacial mais atual e engloba mais regiões do Paraná e início do Mato Grosso do Sul.

Os agrupamentos de baixo risco da análise espaço-temporal também são similares aos agrupamentos da análise espacial de 2019 e englobam principalmente as regiões próximas ao litoral, com algumas diferenças, são elas:

Um agrupamento foi apontado de 2001 a 2011 no interior do Amazonas, ao Sul de Manaus e não existe mais na atual análise espacial, isso significa que a região já foi considerada uma localidade com poucos suicídios e atualmente a taxa vêm se aproximando da média brasileira.

Na região Norte e Nordeste, foi exibido um agrupamento no Norte do Pará e Maranhão envolvendo as capitais e arredores de 1999 a 2009. Diferentemente do que ocorre na análise espacial, a análise espaço-temporal aponta que o agrupamento abrange mais o interior, indicando que atualmente pode haver algum movimento de aumento da taxa no interior. Além disso, neste agrupamento está contida a região ao lado de São Luís, que na análise espacial estava apartada. O agrupamento que englobava Fortaleza e arredores na análise espacial deixa de existir, o que pode significar que esse movimento de baixas taxas de suicídio na localidade é novo.

Na região Nordeste e Sudeste houve um agrupamento que englobou a Bahia e o norte de Minas Gerais de 1997 a 2007 e hoje não existe mais. Esse é um exemplo de região que pode estar aumentando a taxa e se espelhando à taxa média brasileira.

Na região Sudeste, o mesmo agrupamento que engloba a região do Rio de Janeiro continua e o agrupamento da região metropolitana de São Paulo era maior historicamente.

No geral, há alguns agrupamentos e movimentos históricos desses grupos que são destaques para os especialistas da área. O próximo item tem por objetivo apresentar o entendimento de quais são as principais cidades com os maiores crescimentos para se aprofundar na análise, e não somente entender no nível de agrupamentos e regiões.

4.3 ANÁLISE TEMPORAL

Do total de 5.507 municípios, 1.036 municípios são apontados pela regressão de *Joinpoint* com tendência temporal significativa recente (886 aumentando, 94 diminuindo e 56 estáveis). Porém, foram considerados apenas 160 municípios, por conta dos critérios mais conservadores para entendimento se realmente há uma tendência temporal. Desses 160 municípios:

- 80 (50% do total) apontam aumento no número de suicídios.
- 24 (15% do total) apontam diminuição no número de suicídios.
- 56 (35% do total) mantêm números estáveis.

Dentre os municípios identificados com tendências históricas, a

Tabela 4.2 oferece um panorama da presença ou ausência de pontos de inflexão (*joinpoints*) em suas séries temporais. Cautela deve ser exercida ao selecionar um município com uma taxa de suicídio em declínio como ponto de referência, uma vez que tal declínio pode estar relacionado à subnotificação dos casos. Portanto, é recomendável que tanto pesquisadores quanto gestores públicos investiguem o impacto da subnotificação no município em análise. Municípios que não apresentam indícios significativos de subnotificação e que conseguiram reverter uma tendência de aumento para uma de redução das taxas de suicídio assumem relevância, uma vez que podem servir como modelos eficazes no contexto do combate ao suicídio.

	0 joinpoint (tendência hist. contínua)	1 joinpoint (1 ponto inflexão)	2 joinpoints	3 joinpoints	Total
Aumentando	52	21	4	3	80
Diminuindo	15	7	1	1	24
Estável	56	0	0	0	56
Total	123	28	5	4	160

Tabela 4.2 – Resumo da quantidade de municípios que apresentam aumento, diminuição ou estabilidade na taxa de suicídio ajustada por idade e gênero masculino/feminino.

Analisando os municípios que não possuem inflexão em sua série histórica (*joinpoint* = 0), o primeiro município que mais cresce na taxa, percentualmente, é o de Três Lagoas, no Mato Grosso do Sul, divisa com São Paulo, apresentado na Figura 4-27. O segundo é o de Avaré, interior oeste de São Paulo, apresentado na Figura 4-28. O que mais decresce percentualmente a taxa de suicídio é Caarapó, no Mato Grosso do Sul, apresentado na Figura 4-29, e o segundo é São Francisco do Sul, litoral de Santa Catarina, perto da divisa com Paraná, apresentado na Figura 4-30.

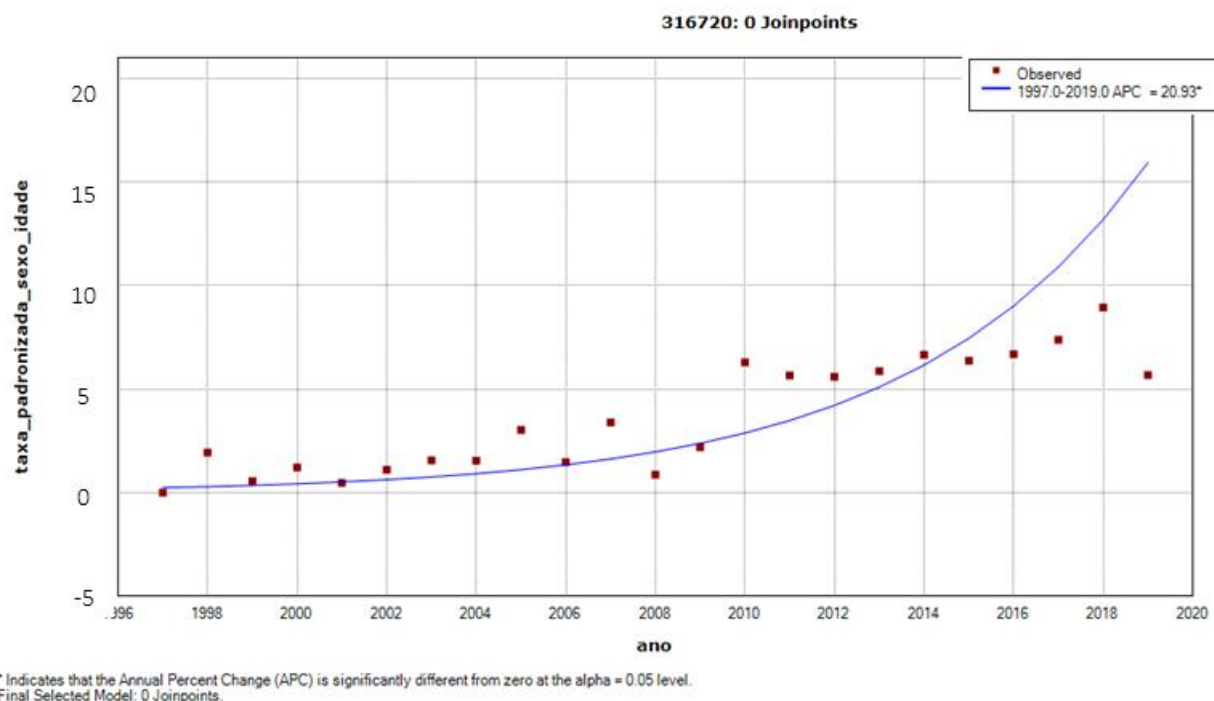


Figura 4-27 – Três Lagoas, município com maior APC com 0 *joinpoint*.

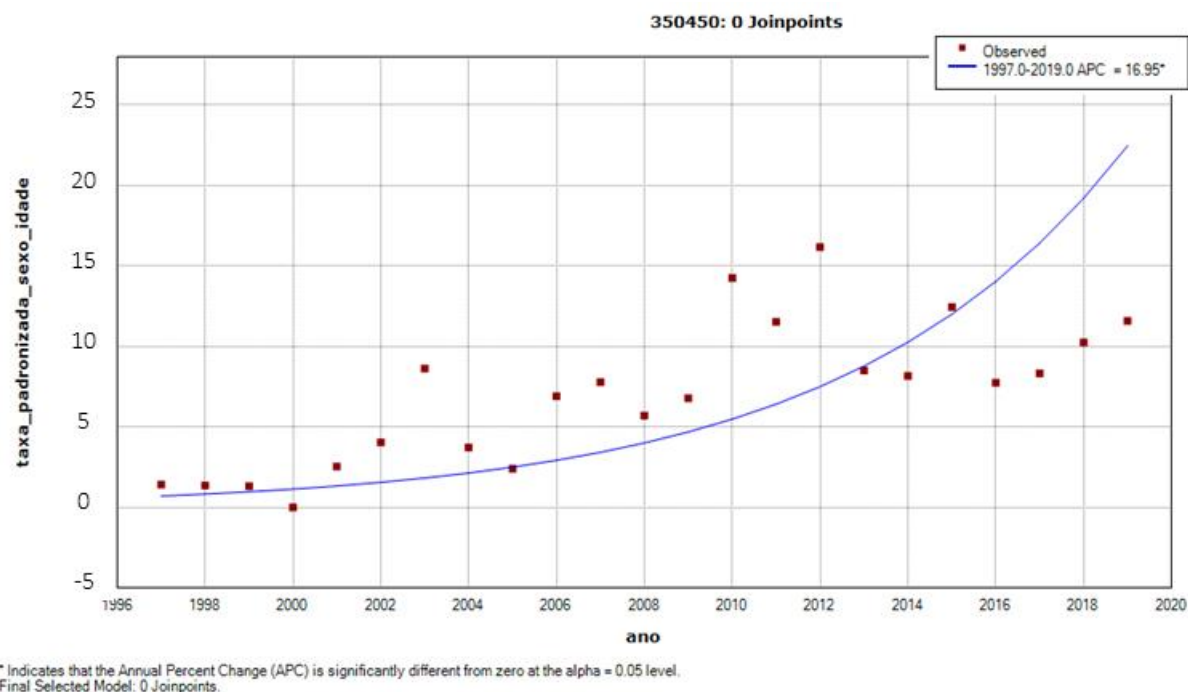


Figura 4-28 – Avaré, segundo município com maior APC com 0 *joinpoint*.

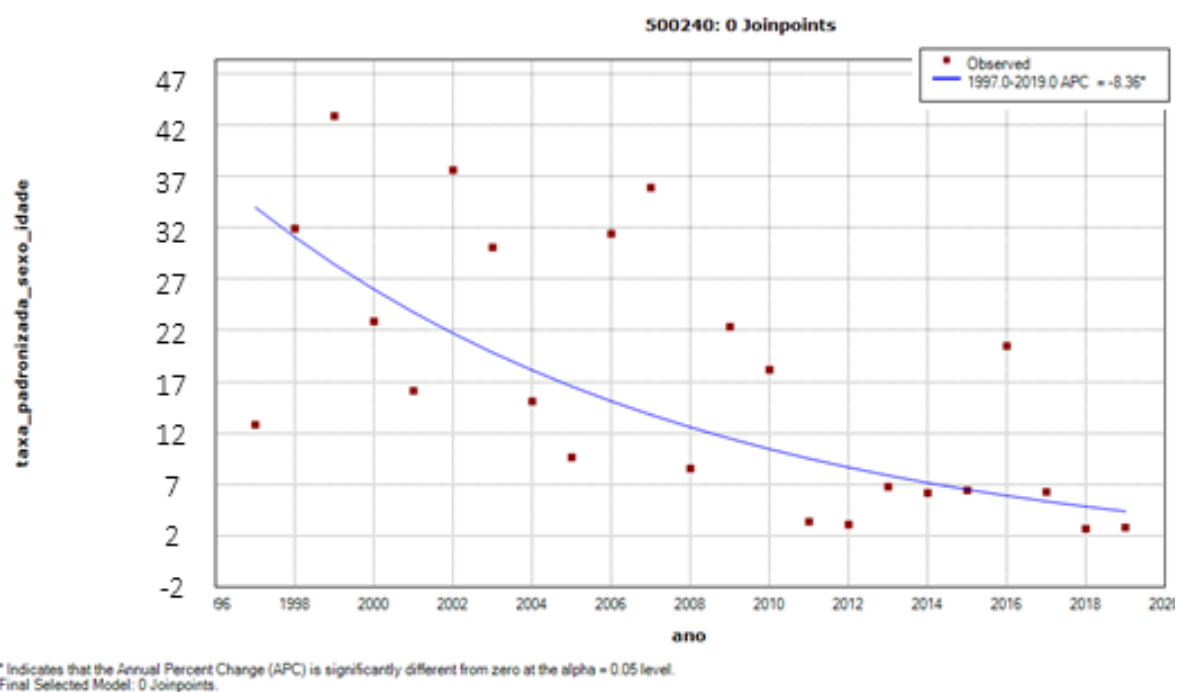


Figura 4-29 – Caarapó, município com menor APC (APC mais negativo) com 0 *joinpoint*.

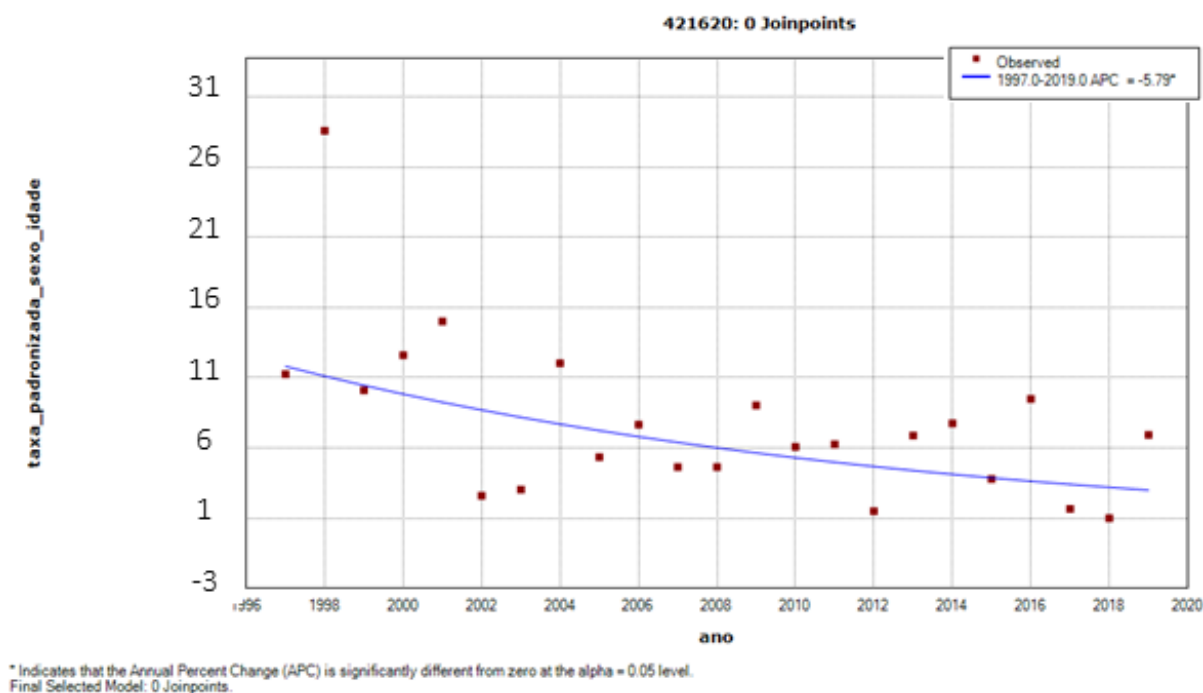
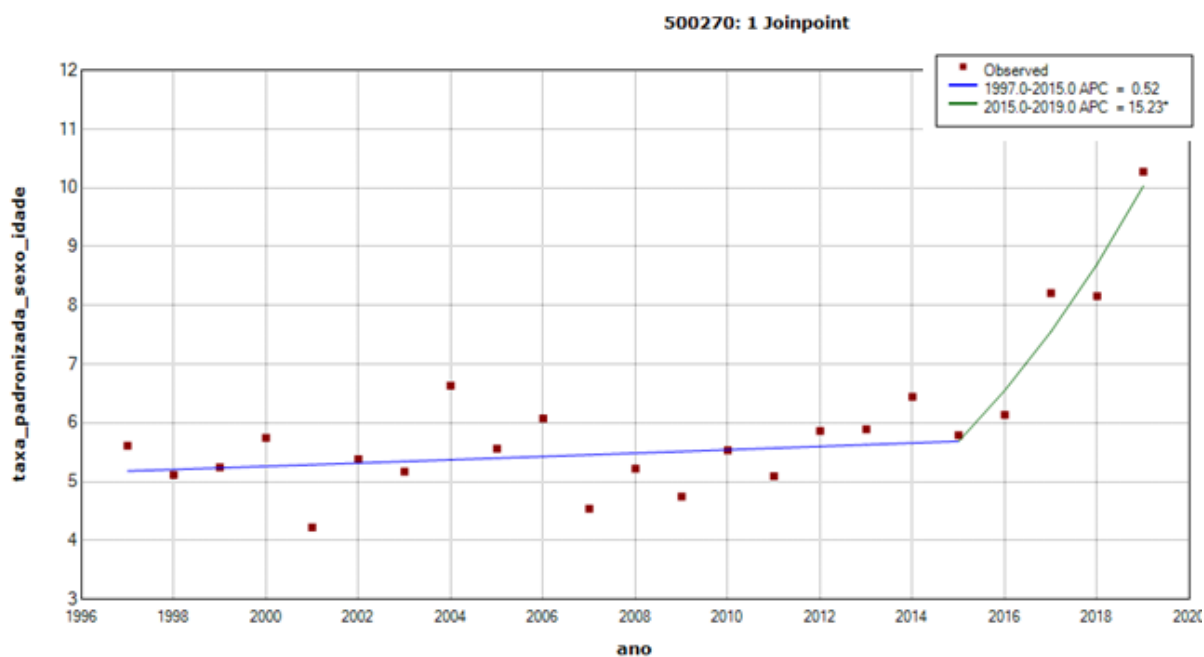


Figura 4-30 – São Francisco do Sul, município com o segundo menor APC com 0 *joinpoint*.

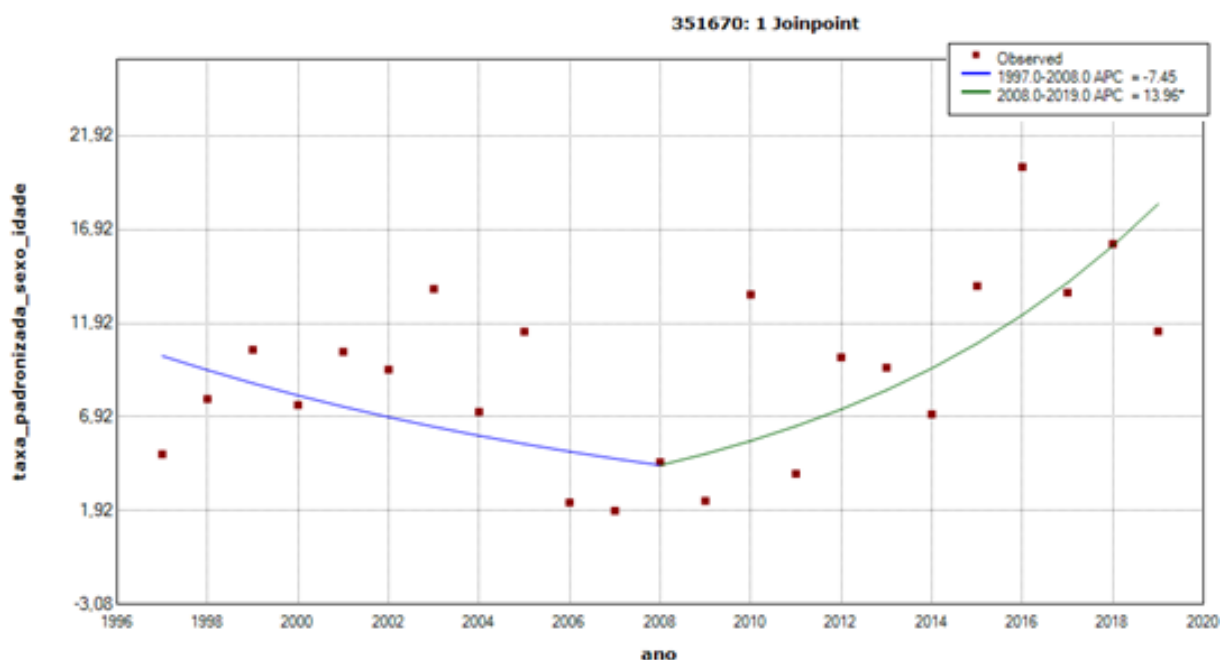
Analisando os municípios que possuem uma inflexão em sua série histórica (*joinpoint* = 1), o primeiro município que mais cresce na taxa, percentualmente, é o de Campo Grande, capital de Mato Grosso do Sul, que apresenta um ponto de inflexão de subida a partir de 2015, apresentado na Figura 4-31. O segundo é Garça, interior oeste de São Paulo, apresentado na Figura 4-32, com ponto de inflexão em 2008.

O município que mais decresce percentualmente a taxa de suicídio com 1 *joinpoint* é Beberibe, no Centro do Ceará, apresentado na Figura 4-33. Observa-se que Beberibe apresentou aumento da taxa até 2001 e depois começou a decrescer. O segundo município é Mossoró, próximo ao litoral de Rio Grande do Norte, perto da divisa com Ceará, apresentado na Figura 4-34. O ponto de inflexão de Mossoró ocorre em 2008, observa-se um aumento da taxa até 2008, seguido de queda. Essas cidades com queda na inflexão são bons exemplos para entender o que possivelmente foi feito de políticas públicas.



* Indicates that the Annual Percent Change (APC) is significantly different from zero at the alpha = 0.05 level.
 Final Selected Model: 1 Joinpoint.

Figura 4-31 – Município de Campo Grande (código município 500270), apresenta um aumento significativo na tendência de aumento de suicídio. Primeiro maior município com APC com 1 *joinpoint*.



* Indicates that the Annual Percent Change (APC) is significantly different from zero at the alpha = 0.05 level.
 Final Selected Model: 1 Joinpoint.

Figura 4-32 – Município de Garça (código município 351670), segundo maior PAC com 1 *joinpoint*

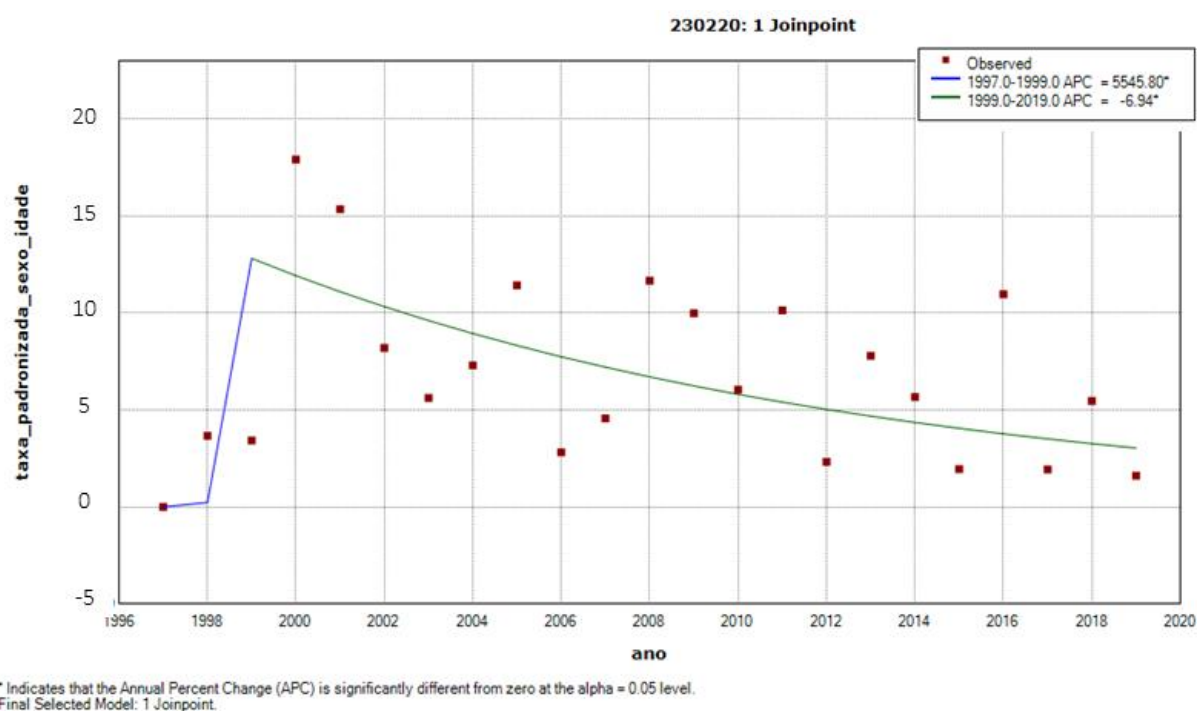


Figura 4-33 – Município de Beberibe. Município com a maior tendência de decrescimento no APC com 1 *joinpoint*.

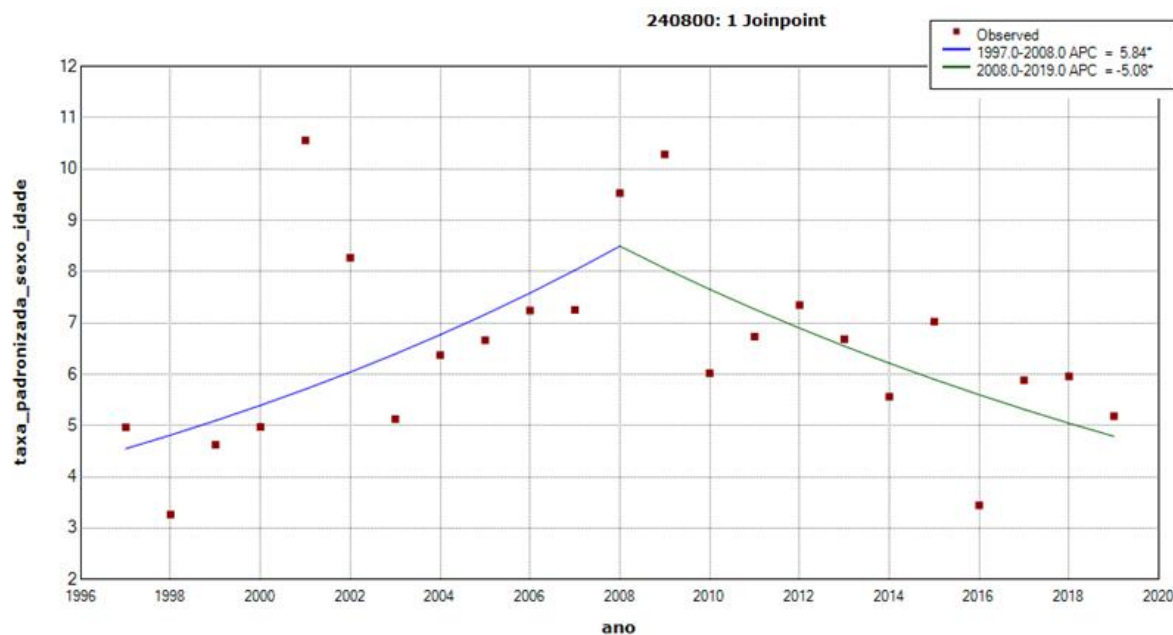


Figura 4-34 – Município de Mossoró. Município com a segunda maior queda de APC com 1 *joinpoint*.

Analisando os municípios que possuem duas inflexões em sua série histórica (*joinpoint* = 2), o primeiro que mais cresce na taxa percentual atualmente é o de Cascavel, no interior oeste do estado do Paraná. O município, a partir de 2009, vem crescendo continuamente sua taxa. O município com maior decréscimo da taxa é São Paulo, capital do estado de São Paulo. Observa-se que o município começa a decrescer a taxa em 2014, alcançando taxas mínimas em seu histórico.

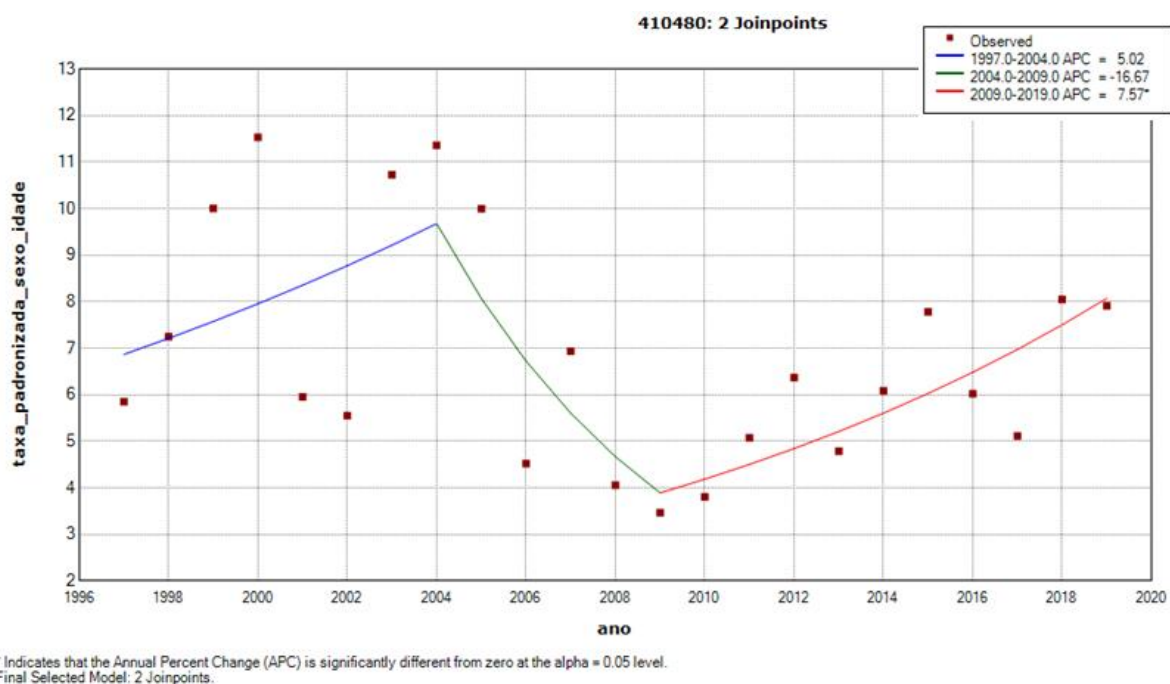


Figura 4-35 – Município Cascavel, com maior APC com 2 *joinpoints*

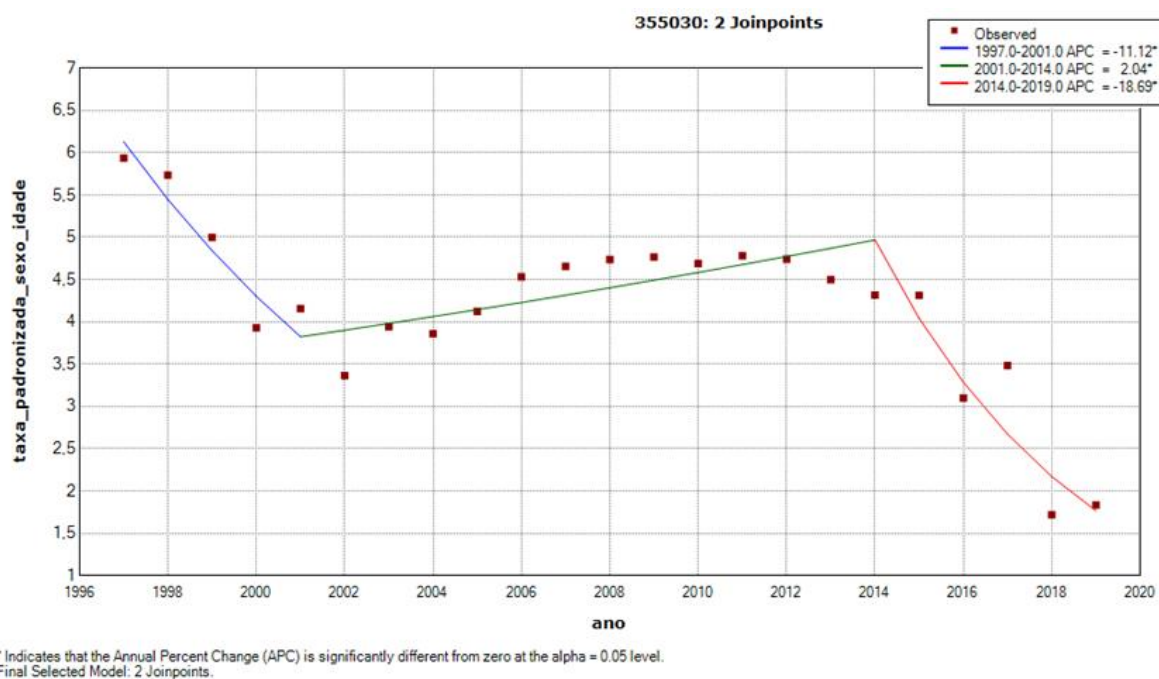


Figura 4-36 – Município São Paulo com menor APC com 2 *joinpoints*

Analisando os municípios que possuem três inflexões em sua série histórica (*joinpoint* = 3), destaque para a cidade de Santana de Parnaíba, pertencente à região metropolitana de São Paulo, que apresenta tendência de crescimento desde 2008 e já decresceu em alguns períodos. Destaque para o decrescimento da taxa em Itapeverica da Serra, região metropolitana de São Paulo, que historicamente vem decrescendo sua taxa, porém entre 2006 e 2010 apresentou um aumento, voltando a decrescer em 2011.

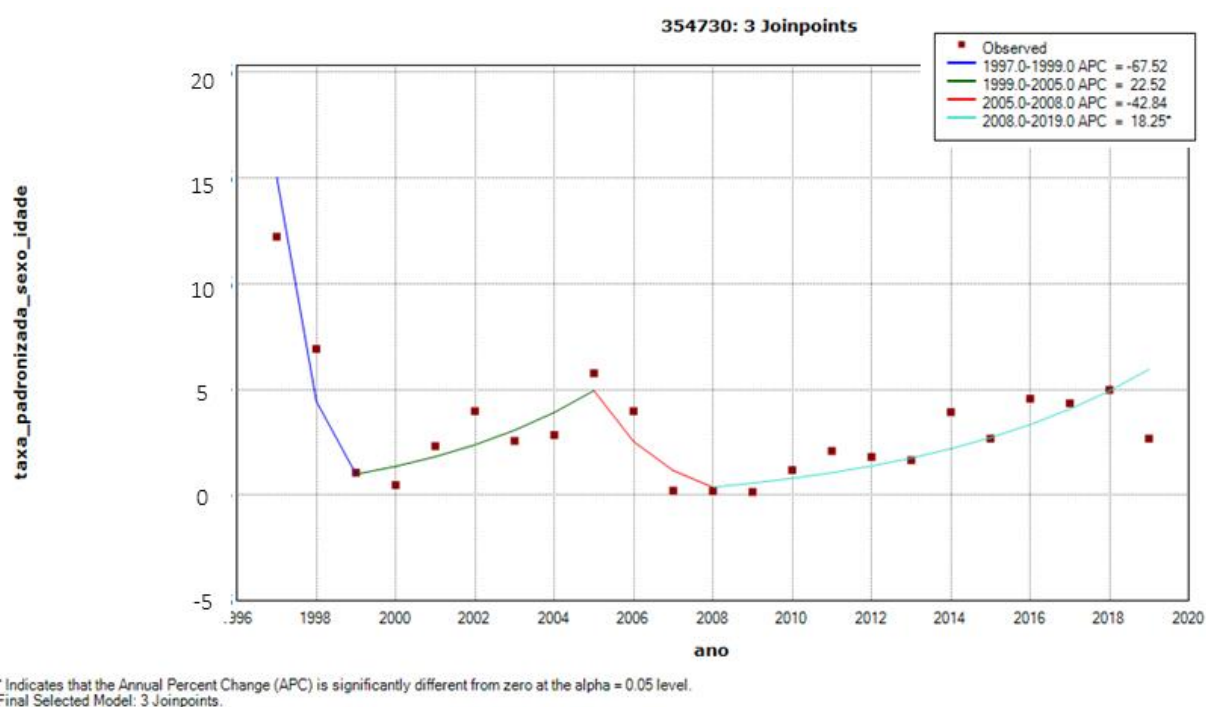


Figura 4-37 – Município Santana de Parnaíba com tendência de crescimento com 3 *joinpoints*

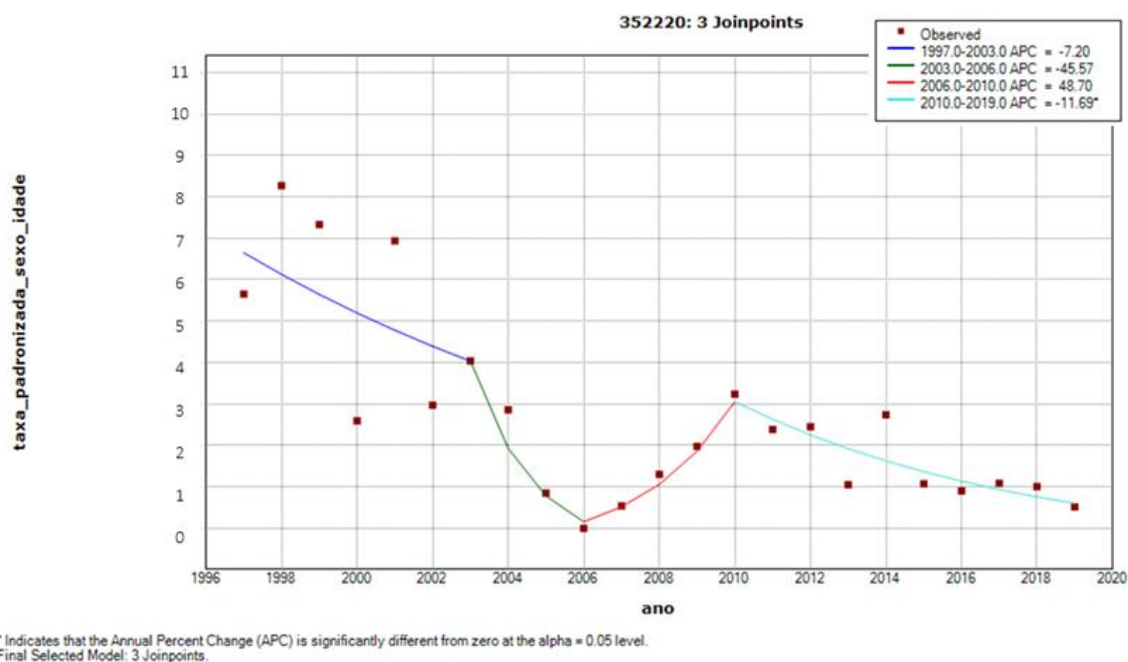


Figura 4-38 – Município Itapecerica da Serra com tendência de queda com 3 *joinpoints*

Por meio dos painéis em Dada (2022) é possível avaliar a tendência histórica de cada cidade para, por exemplo, o gestor público poder analisar o efeito dos seus estudos e ações.

4.4 ANÁLISE MULTIVARIADA

Para entender se há correlação entre os tipos de urbanização/ ruralização e os agrupamentos de alto e baixo risco, realizou-se o teste de qui-quadrado, e verificou-se que não há diferença entre as proporções das classificações urbanas com os agrupamentos de alto e baixo risco. O p-valor encontrado foi de 7%, aceitando a hipótese nula de que ambas as proporções são iguais. Para se rejeitar a hipótese, estava sendo considerado um p-valor menor que 5%. A Figura 4-39 representa as proporções encontradas entre os agrupamentos.

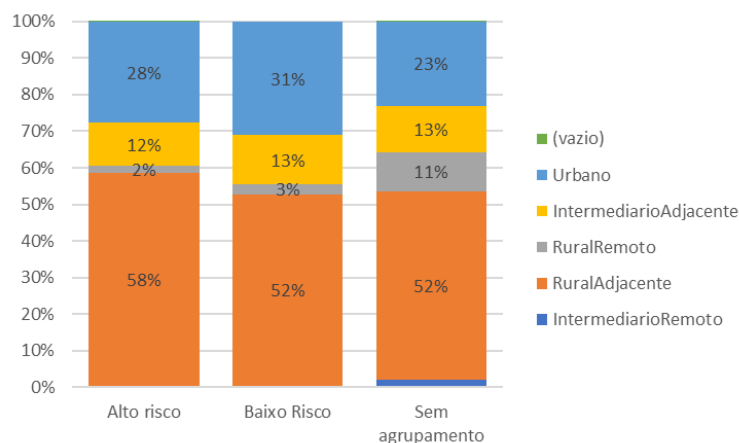


Figura 4-39 – Proporções entre os tipos de urbanização e os agrupamentos de alto e baixo risco.

Conforme visto nos subitens anteriores, não há um padrão específico para os suicídios no Brasil em geral, possivelmente cada região apresenta fatores de risco do suicídio que são mais específicos do que se é vivido naquela localidade. Por isso, como trabalho futuro, sugere-se realizar esse mesmo teste estatístico para cada unidade federativa brasileira considerando os municípios de baixo e alto risco encontrados em cada um dos estados, ou aglomerado de estados caso não haja municípios suficientes em cada um dos agrupamentos e a amostragem fique baixa (que pode gerar anomalias).

5 CONCLUSÕES E PESQUISAS FUTURAS

A pesquisa desenvolvida incluiu uma revisão bibliográfica sobre quatro assuntos. São eles: **(1)** suicidologia, **(2)** bases de dados que podem ser utilizadas para se estudar características de saúde e socio-demografia no Brasil, **(3)** estudo espaço-temporal de um evento na área de ciência de dados e, para finalizar, **(4)** estudo espaço-temporal do suicídio.

Observou-se que estudos espaço-temporais possuem diversas abordagens estatísticas e computacionais para problemas diferentes (saúde, clima, crime etc.). Constatou-se, também, que normalmente, no estudo do suicídio, se faz uma análise descritiva, seguida de uma análise espacial, uma análise temporal e, por fim, uma análise multivariada.

Analisando a taxa de suicídio obtida pela OMS (6,41/100.000 habitantes em 2019) e a pelo estudo (6,07/100.000 habitantes), que são diferentes em anos de projeção da população, destaca-se a importância de haver Censo no país regularmente, e a importância de se disseminar os mesmos cálculos de projeção da população entre as organizações. Dessa forma, é possível comparar o país com o restante do mundo e avaliar possíveis políticas públicas a serem tomadas, que foram adotadas em países com mesmos aspectos culturais que o nosso.

Apesar de as taxas serem diferentes, conclui-se que a taxa de suicídio está aumentando historicamente no Brasil, entre homens e mulheres e todas as faixas etárias, com exceção de pessoas com mais de 75 anos. No Brasil, assim como no mundo, a taxa de suicídio nas mulheres (2,59/100.000 habitantes) é menor do que entre homens (8,88/100.000 habitantes) e a taxa de suicídio é maior quanto maior a faixa etária.

Porém, entre os mais jovens, as taxas vêm crescendo mais, independente do gênero masculino/feminino, com destaque para a faixa etária de 15 a 24 anos, com aumento de 150% em um intervalo de 24 anos.

Como foi verificada a diferença significativa da taxa de suicídio entre o gênero masculino/feminino e faixa de idade, reforça-se a importância de o cálculo da taxa ser

padronizado por gênero masculino/feminino e idade, e não somente por idade ou com as taxas brutas. A taxa padronizada por gênero masculino/feminino e idade não trata as anomalias que podem ocorrer em cidades com populações pequenas, ocorrendo intermitência na taxa de suicídio em alguns locais e valores altos anômalos nos mesmos lugares. Para isso, neste estudo, esses locais são desconsiderados em determinadas análises para se evitar conclusões incorretas. Como pesquisa futura, recomenda-se um estudo de comparação dos resultados obtidos pela taxa padronizada por idade e gênero masculino/feminino e a taxa com suavização espacial para se entender como as diferenças de metodologia podem impactar no resultado.

Ao se analisar as grandes cidades que podem impactar mais na taxa crescente de suicídio no Brasil por serem mais populosas e influenciarem mais o cálculo ponderado, entende-se que o suicídio está mais pulverizado do que a concentração da população e que não há alguma cidade anômala que seja responsável por grande parte do resultado do Brasil. Destaque para as cidades de Teresina (PI), Campo Grande (MS) e Porto Alegre (RS) que possuem taxas maiores que 8 e para o crescimento acima de 170% da taxa para as cidades de Campinas (SP), João Pessoa (PB), Duque de Caxias (RJ), Teresina (PI) e São Bernardo do Campo (SP).

Analisando-se espacialmente, entende-se que há agrupamentos de alto risco de suicídio em todas as regiões do Brasil. Destaque para os agrupamentos de alto risco no Norte, divisa com Peru e Colômbia, no semiárido interior do Ceará e Piauí, na região central do Brasil, em uma pequena região no Sul de Minas e na região Sul.

Nos agrupamentos do Norte e Centro-Oeste, onde se observou alto risco de suicídio, algumas referências sugerem que uma das possíveis razões seria devido à população indígena estar em maiores proporções nesses locais. Há diversos fatores que levam a população indígena a cometer mais o ato que o restante da população, como a desagregação cultural (PALMA; SANTOS; IGNOTTI, 2020; ORELLANA et al., 2016).

No contexto do agrupamento geográfico no Sul do Brasil, ao considerar exclusivamente os fatores relacionados à precipitação, há indicações de que a incidência elevada de suicídio está associada a aglomerações geográficas

caracterizadas por privações econômicas e infraestrutura deficiente em ambientes rurais. Adicionalmente, existem indícios de que tais padrões de comportamento social, influenciados pela colonização europeia, desempenham um papel relevante nesse cenário (ALARCÃO et al., 2019; PALMA; SANTOS; IGNOTTI, 2020). Por outro lado, é importante notar que, na mesma região Sul, existem áreas com Produto Interno Bruto (PIB) elevado que também apresentam altas taxas de suicídio. Essa complexidade sugere, em primeiro lugar, que a ocorrência do suicídio não pode ser completamente explicada apenas pelos fatores relacionados à precipitação e, em segundo lugar, que existe uma necessidade de aprofundamento na pesquisa da região Sul.

Não foram encontrados estudos específicos sobre o agrupamento do Sul de Minas e a região do semiárido Nordeste que envolve o Ceará e o Piauí especificamente. Sugere-se que cada área seja estudada separadamente porque, possivelmente, há razões diferentes para cada área apresentar alto risco, ainda mais considerando-se que não houve diferenciação na taxa de ruralidade entre agrupamentos de alto e baixo risco.

Há também agrupamentos de baixo risco que são possíveis exemplos a serem seguidos por gestores públicos ou podem significar casos de subnotificação. São eles: parte do litoral do Nordeste, região metropolitana do Rio de Janeiro e São Paulo. Os agrupamentos de baixo risco estão nas regiões litorâneas e populosas do Brasil. Há estudos para se entender as motivações do suicídio nessas regiões (BANDO et al., 2012; SANTOS et al., 2018), porém não foram encontrados estudos que visam entender se há algum fator que contribui para essas cidades apresentarem as menores taxas de suicídio do Brasil ou tendências a subnotificação. Recomenda-se este estudo futuro também. Adicionalmente, a fim de sustentar a alegação de que esses aglomerados efetivamente representam áreas de baixo risco, é aconselhável realizar uma análise prévia da subnotificação de casos.

Pelos agrupamentos de baixo risco englobarem as regiões metropolitanas, pode-se gerar alguma expectativa de que, quanto maior o PIB *per capita*, menor o suicídio em todos os territórios no Brasil. Porém, essa afirmação não é verdadeira porque há agrupamentos de alto risco que englobam regiões de alto PIB *per capita*,

como a região Sul. Dessa forma, reforça-se a necessidade de se estudar cada área isoladamente porque há indícios de que as razões são diversas.

Analisando-se os agrupamentos espaço-temporais, entende-se que os mesmos agrupamentos espaciais estão na análise temporal, com a adição de uma região de baixo risco no interior do Amazonas de 2001 a 2011.

Na análise temporal, observa-se, primeiramente, que cada município possui sua tendência histórica específica, ou seja, não é possível observar um agrupamento das séries históricas entre os municípios.

Os municípios de Três Lagoas, Avaré, Campo Grande, Garça, Cascavel e Santana de Parnaíba apresentam historicamente as maiores tendências de crescimento do suicídio, sendo um ponto de atenção aos seus gestores públicos.

Além dos municípios destacados nesta conclusão como pontos de atenção ou como exemplos, no painel em Dada (2022) está disponível uma ferramenta para ser possível se aprofundar na análise de cada cidade do Brasil de uma forma prática.

Há oportunidade de aprofundamento técnico em cada uma das etapas da análise espaço-temporal do suicídio, com comparações reais entre as abordagens estatísticas e computacionais e a escolha da mais adequada a partir de um *ensemble* do melhor desempenho.

Durante o processo de coleta de dados, é plausível que a subnotificação de casos de suicídio possa ocorrer devido a diversas influências, tais como aquelas de natureza religiosa, social, cultural, política e econômica, conforme apontado por PALMA, SANTOS e IGNOTTI (2020). Uma abordagem para investigações futuras consiste na aplicação de técnicas da área de ciência de dados, a fim de identificar possíveis indícios de subnotificação de casos de suicídio. Esse procedimento pode envolver uma análise de frequência das causas de morte em cada cidade, considerando fatores religiosos e sociodemográficos específicos para cada localidade em estudo.

Além disso, outra oportunidade de pesquisa a ser considerada é a inclusão, além dos casos de suicídio, das mortes com intenção indeterminada no escopo da

investigação. Essa ampliação do enfoque permitiria a exploração de eventuais erros de classificação das causas de morte, seguindo a abordagem adotada em outros países, como mencionado por MCDONALD et al. (2021).

Outra oportunidade de continuidade da pesquisa é na realização do estudo para as cidades que foram excluídas da análise (0,01%) por falta de dados de geolocalização ou de histórico de ao menos de 18 anos. Para o administrador público e especialista do assunto, cada município é relevante e há planos de ação individualizado para cada cidade.

Em virtude dos fatos mencionados, conclui-se que, no Brasil, de maneira geral, a taxa de suicídio vem aumentando e não há uma razão geral para isso. Cada localidade deverá ser analisada, comparando-se com cidades similares que reverteram a situação.

Espera-se, com a conclusão desta pesquisa, que a complexidade de acesso, falta de padronização e confiabilidade nos dados de suicídio do Brasil sejam minimizados com essa pesquisa. Além disso, com essa pesquisa dedicada à mineração de dados (*data-mining*), deseja-se oferecer uma metodologia e uma ferramenta, através desta dissertação e dos painéis em Dada (2022), que possibilitem aos usuários, especialmente pesquisadores e gestores públicos interessados na área, fazer análises de um banco de dados compilado, podendo tanto tirar rápidas conclusões, quanto se aprofundar nos motivos incentivadores do ato, em cada município e nos agrupamentos de alto risco destacado no estudo.

Por meio dessa ferramenta, os usuários têm a capacidade de escolher uma ou várias cidades e realizar e visualizar no mapa uma comparação da incidência de suicídios, considerando variáveis como gênero, faixa etária e taxa de urbanização, ao longo do período de 1996 a 2019. Inicialmente, a criação dessa ferramenta abrangendo todo o território brasileiro demandou um total de 480 horas de desenvolvimento. Com o uso da ferramenta, os gestores podem obter os resultados desejados em questão de segundos.

Ressaltam-se a complexidade para obtenção dos dados apresentada no item 3.1 DADOS e as inconsistências encontradas nas bases de dados públicas existentes,

como a falta de municípios em alguns anos ou a falta de geolocalização de algumas cidades. Esses fatores evidenciam como ainda há informações públicas que poderiam ser mais aderentes aos princípios de dados abertos, ou dados FAIR (WILKINSON et al., 2016), ou seja, essas informações poderiam ser mais facilmente localizáveis, acessíveis, interoperáveis e reutilizáveis, além de mais confiáveis. Por ser a saúde uma questão de Estado, e por muitos dados serem necessários à gestão da saúde são gerados e/ou geridos por órgãos públicos, é fundamental que seja dada maior importância a eles. Isso inclui políticas e diretrizes sobre dados da área da saúde e deve incluir a política de dados abertos, evidentemente com as restrições aplicáveis relativas a dados pessoais, por exemplo.

Duas iniciativas em curso merecem menção e incentivo: GO Fair (GO FAIR, 2023) e *Open Government Partnership* (OGP) (OGP, 2023). GO Fair é um movimento internacional que incentiva a adoção e a prática dos princípios FAIR. O Brasil faz parte desse movimento por meio da GO FAIR Brasil (GO FAIR BRASIL, 2021) e entre as redes temáticas de implementação, há a GO FAIR Brasil Saúde. Já a *Open Government Partnership*, é uma iniciativa da qual o Brasil é membro fundador, e que visa difundir práticas governamentais para transparência, acesso à informação e participação social. No Brasil ela é liderada pela Controladoria Geral da União, e se desenvolve na forma de planos de ação que englobam diferentes áreas, incluindo a saúde. Nos anos recentes, por exemplo, os planos incluíram padrões e diretrizes voltados à interoperabilidade e usabilidade de sistemas relacionados ao Sistema Nacional de Vigilância Sanitária (OGP, 2023). Considerando a experiência no desenvolvimento desta dissertação, as dificuldades com diversas questões ligadas aos dados, essas iniciativas são de muita importância e espera-se que tenham seguimento e cresçam.

O presente estudo apresenta algumas limitações, incluindo a subnotificação dos casos de suicídio, a exclusão de cidades devido à ausência de técnicas de suavização espacial e a natureza de pesquisa ecológica, não definindo relações de causalidade. Ademais da restrição previamente aludida, ainda está pendente a determinação do método de disponibilização da pesquisa às Secretarias Municipais e Estaduais de Saúde.

A etiologia do suicídio é intrincada, envolvendo desafios no entendimento dos fatores da predisposição ao ato, que incluem elementos como transtornos psicológicos e hereditariedade. Além disso, a análise dos fatores de precipitação ao ato, que abarcam influências externas, é uma tarefa complexa e multidisciplinar. Nesse contexto, a utilização da computação em conjunto com métodos estatísticos, que culminam na aplicação de técnicas de inteligência artificial e aprendizado estatístico, emerge como uma abordagem promissora para aprofundar o entendimento desse fenômeno complexo. É saliente a relevância das ferramentas e metodologias computacionais e estatísticas na condução deste estudo, e a sua crescente importância nas tomadas de decisão no contexto de áreas multidisciplinares.

Finalmente, cumpre lembrar que, com o agravamento da crise global da Saúde Planetária, fatores desencadeadores de suicídios tendem a aumentar sua incidência no mundo, o que torna ainda mais urgente entender-se e agir para mitigar o problema. Uma perspectiva de pesquisa futura envolve a investigação dos impactos da crise global da Saúde Planetária na incidência de suicídio em diferentes cidades brasileiras. Simultaneamente, é urgente trabalhar para reverter o impacto humano nos sistemas naturais, que demandará mudanças de comportamento humano, como bem expressa a Declaração de São Paulo sobre Saúde Planetária (MYERS, PIVOR, SARAIVA, 2022).

REFERÊNCIAS BIBLIOGRÁFICAS

ABUABARA, A.; ABUABARA, A.; TONCHUK, C. A. L. Comparative analysis of death by suicide in Brazil and in the United States: descriptive, cross-sectional time series study. **Sao Paulo Medical Journal**, v. 135, n. 2, p. 150–156, 2017.

AHMAD, O. B.; BOSCHI-PINTO, C.; LOPEZ, A. D. Age standardization of rates: a new WHO standard. **GPE Discussion Paper Series**, n. 31, p. 1–14, 2001.

ALARCÃO, A. C. et al. Suicide mortality among youth in southern Brazil: a spatiotemporal evaluation of socioeconomic vulnerability. **Brazilian Journal of Psychiatry**, v. 42, n. 1, p. 46–53, 2019.

ANACONDA. Disponível em: <<https://www.anaconda.com/>>. Acesso em 10 fev. 2023.

ARCGIS ONLINE. Disponível em: <https://www.arcgis.com/index.html>. Acesso em: 25 out. 2020.

ATLURI, G.; KARPATNE, A.; KUMAR, V. Spatio-temporal data mining: A survey of problems and methods. **ACM Computing Surveys**, v. 51, n. 4, p. 1–37, 2018.

BANDO, D. H. et al. Suicide rates and income in São Paulo and Brazil: A temporal and spatial epidemiologic analysis from 1996 to 2008. **BMC Psychiatry**, v. 12, 2012.

BARROZO, L.V. Currículo do sistema currículo Lattes. [São Paulo], 02 out. 2023. Disponível em: <<http://lattes.cnpq.br/0204977512052749>>. Acesso em: 21 out. 2023.

BERTOLETE, J. M.; FLEISCHMANN, A. Suicidal behavior prevention: WHO perspectives on research. **American Journal of Medical Genetics - Seminars in Medical Genetics**, v. 133 C, n. 1, p. 8–12, 2005.

BIRANT, D.; KUT, A. ST-DBSCAN: An algorithm for clustering spatial-temporal data. **Data and Knowledge Engineering**, v. 60, n. 1, p. 208–221, 2007.

BRASIL. Lei nº 5.172, de 25 de outubro de 1966. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/l5172compilado.htm>. Acesso em: 09 ago. 2022.

BRUNSDON, C.; FOTHERINGHAM, S.; CHARLTON, M. Geographically weighted regression Φ modelling spatial non-stationarity. **Journal of the Royal Statistical Society Series D: The Statistician**, p. 431–443, 1998.

CHIU, B.; KEOGH, E.; LONARDI, S. Probabilistic discovery of time series motifs. In: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING. **Proceedings** [...]. Washington, D.C., US, 2003, p. 493–498.

DADA, S.G. **Suicide in Brazil. Tableau**, 2022. Disponível em <https://public.tableau.com/app/profile/sheila.genesine.dada/viz/painel_suicidio/Viso geral1>. Acesso em 08 ago. 2022.

DADA, S. G. "Suicides in Brazil from 1996 to 2019 per city, sex and age", **Harvard Dataverse**, V1, 2023. <https://doi.org/10.7910/DVN/DW1MOO>.

DANTAS, A.P. et al. Analyss of suicide mortality in Brazil: spatial distribution and socioeconomic context. **Braz J Psychiatry**. 40:12-8.38, 2018.

DATASUS. **População residente** – Brasil. Disponível em: <<http://tabnet.datasus.gov.br/cgi/deftohtm.exe?ibge/cnv/popbr.def>>. Acesso em: 09 ago. 2022.

DATASUS. **População residente** - estudo de estimativas populacionais para os municípios brasileiros, desagregadas por sexo e Idade, 2000-2021. Disponível em: <<http://tabnet.datasus.gov.br/cgi/POPSVS/NT-POPULACAO-RESIDENTE-2000-2021.PDF>>. Acesso em: 09 ago. 2022a.

DATASUS. **População residente** – Notas técnicas. Disponível em: <<http://tabnet.datasus.gov.br/cgi/lbge//popdescr.htm>>. Acesso em: 09 ago. 2022b.

DATASUS. **SIM/CID-10**, 2020b. Disponível em: <<ftp://ftp.datasus.gov.br/dissemin/publicos/SIM/CID10/DORES/>>. Acesso em: 05 mai. 2020.

DATASUS. **Sobre o datasus**, c2019a. Disponível em: <<https://datasus.saude.gov.br/sobre-o-datasus/>>. Acesso em: 19 abr. 2020.

DATASUS. **Tabnet**, c2019b. Disponível em: <<https://datasus.saude.gov.br/informacoes-de-saude-tabnet/>>. Acesso em: 19 abr. 2020.

DATASUS. **Tutorial TABNET 2020**, c2020. Disponível em: <<https://datasus.saude.gov.br/wp-content/uploads/2020/02/Tutorial-TABNET-2020.pdf>>. Acesso em: 19 abr. 2020.

DATAVERSE. Disponível em: <<https://dataverse.org/>>. Acesso em 09 fev. 2023.

DBASE. Disponível em: <<https://www.dbase.com/>>. Acesso em: 04 fev. 2023.

DURKHEIM, E. **O suicídio**: estudo de sociologia. São Paulo: Martins Fontes, 2000.

EGGER M.; SMITH G.D.; ALTMAN D. **Systematic reviews in health care**: meta-analysis in context. 2. ed. Londres: BMJ Publishing Group, 2008. ISBN:9780470693926.

EL-SAPPAGH, S. H. A.; HENDAWI, A. M. A.; EL BASTAWISSY, A. H. A proposed model for data warehouse ETL processes. **Journal of King Saud University - Computer and Information Sciences**, v. 23, n. 2, p. 91–104, 2011.

EXETER, D.J.; BOYLE, P.J.. Does young adult suicide cluster geographically in Scotland? **Journal of Epidemiology and Community Health**, 61:731-736, 2007.

FAZEL, S.; RUNESON, B. Suicide. **The New England journal of medicine**, v. 382, n. 3, p. 266–274, 2020.

FONTANELLA, C. A. et al. Mapping suicide mortality in Ohio: A spatial epidemiological analysis of suicide clusters and area level correlates. **Preventive Medicine**, v. 106, n. May 2017, p. 177–184, 2018.

FTP DATASUS. Disponível em: <ftp://ftp.datasus.gov.br/>. Acesso em: 25 out. 2020.

GITHUB. Disponível em: <https://github.com/>. Acesso em: 25 out. 2020.

GO FAIR. Disponível em: <https://www.go-fair.org/>. Acesso em 03 abr. 2023.

GO FAIR BRASIL. Início, 2021. Disponível em: <https://www.go-fair-brasil.org/>. Acesso em 03 abr. 2023.

HAYES, K. et al. Climate change and mental health: Risks, impacts and priority actions. **International Journal of Mental Health Systems**, v. 12, n. 1, p. 1–12, 2018.

IBGE - DIRETORIA DE GEOCIÊNCIAS. **Base cartográfica contínua do Brasil – Documentação técnica geral**, 2016. Disponível em: https://geoftp.ibge.gov.br/cartas_e_mapas/bases_cartograficas_continuas/bcim/ver_sao2016/informacoes_tecnicas/documentacao_tecnica/DocTecnica_BCIM_VOL_II_Anexos_03nov16.pdf. Acesso em: 11 ago. 2022.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Censo 2010 – Glossário**, 2022. Disponível em: <https://Censo2010.ibge.gov.br/materiais/guia-do-Censo/glossario.html>. Acesso em: 09 ago. 2022.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Censo Demográfico – O que é**, a. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9662-Censo-demografico-2010.html?=&t=o-que-e>. Acesso em: 07 jun. 2020.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Censo Demográfico – Microdados**, b. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9662-Censo-demografico-2010.html?=&t=microdados>. Acesso em: 07 jun. 2020.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Classificação e Caracterização dos Espaços Rurais e Urbanos do Brasil**. Disponível em: <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/15790-classificacao-e-caracterizacao-dos-espacos-rurais-e-urbanos-do-brasil.html?=&t=acesso-ao-produto>. Acesso em: 01 jun. 2022a.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Bases cartográficas contínuas – Brasil**. Disponível em: <<https://www.ibge.gov.br/geociencias/cartas-e-mapas/bases-cartograficas-continuas/15759-brasil.html?=&t=downloads>>. Acesso em: 11 ago. 2022c

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Localidades**. Disponível em: <https://geofpt.ibge.gov.br/organizacao_do_territorio/estrutura_territorial/localidades/>. Acesso em: 09 ago. 2022b.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Estimativas da população**. Disponível em: <<https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao>>. Acesso em: 04 fev. 2023

IEEE EXPLORE. Disponível em: <https://ieeexplore.ieee.org/Xplore/home.jsp>. Acesso em: 04 ago. 2022.

JAEN-VARAS, D. et al. The association between adolescent suicide rates and socioeconomic indicators in Brazil: a 10-year retrospective ecological study. **Brazilian Journal of Psychiatry**, 2019.

JONES, P. et al. Identifying Probable Suicide Clusters in Wales Using National Mortality Data. **PLoS One**, 8:e71713, 2013

JUPYTER. Disponível em: <<https://jupyter.org/>>. Acesso em 10 fev. 2023.

KELEJIAN, H. H.; PRUCHA, I. R. A generalized moments estimator for the autoregressive parameter in a spatial model. **International Economic Review**, v. 40, n. 2, p. 509–533, 1999.

KHANA D. et. al. A bayesian spatial and temporal modeling approach to mapping geographic variation in mortality rates for subnational areas with r-inla. **Journal of Data Science**. Jan;16(1): p. 147-182. PMID: 29520299; PMCID: PMC5839164, 2018.

KIM H. J. et al. Permutation tests for joinpoint regression with applications to cancer rates. **Statistics in Medicine**, 19:335-351, 2000.

KULLDORFF, M. A spatial scan statistic. **Communications in Statistics - Theory and Methods**, v. 26, n. 6, p. 1481–1496, 1997.

KULLDORFF, M. et al. A space-time permutation scan statistic for disease outbreak detection. **PLoS Medicine**, v. 2, n. 3, p. 0216–0224, 2005.

KULLDORFF, M. Prospective time periodic geographical disease surveillance using a scan statistic. **Journal of the Royal Statistical Society. Series A: Statistics in Society**, v. 164, n. 1, p. 61–72, 2001.

KRUG, E. G. et al. **World report on violence and health**. Geneva: World Health Organization, 2002.

LOPES, H. E. G.; GOSLING, M. S. Cluster Analysis in Practice: Dealing with Outliers in Managerial Research. **RAC - Revista de Administração Contemporânea (Journal of Contemporary Administration)**, v. 25, n.1, e200081, 2021.

LOZANO, A. et al. Spatial-temporal causal modeling for climate change attribution. In.: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING. **Proceedings** [...]. Paris, France, 2009, p. 587–595.

LUO, Q. et al. Spatio-temporal Granger causality: A new framework. **NeuroImage**, v. 79, p. 241–263, 2013.

MAIA, A. G. **Econometria: Conceitos e Aplicações**. Saint Paul Editora, cap. 12, 2018.

MCDONALD, K. et al. Trends in method-specific suicide in Brazil from 2000 to 2017. **Social Psychiatry and Psychiatric Epidemiology**. v. 56, p. 1779–1790, 2021. <https://doi.org/10.1007/s00127-021-02060-6>.

MEDRONHO, R.A.; BLOCH, K. V.; WERNECK, G. L. **Epidemiologia**. São Paulo: Editora Atheneu; 2006.

MENG, J. et al. Mining Motifs from Human Motion. **Eurographics 2008 Short Papers**, p. 1–4, 2008.

MESOUDI, A. The cultural dynamics of copycat suicide. **PLoS One**, 4:e7252, 2009.

MINNEN, D. et al. Discovering multivariate motifs using subsequence density estimation and greedy mixture learning. In: NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE. **Proceedings** [...]. Vancouver, British Columbia, Canada, 2007, v. 1, p. 615–620.

MYERS, S. S.; PIVOR, J. I.; SARAIVA, A. M. The São Paulo Declaration on Planetary Health. **The Lancet**, v. 398, n. 10308, p. 1299, out. 2021.

NCI – SRP - NATIONAL CANCER INSTITUTE - SURVEILLANCE RESEARCH PROGRAM, STATISTICAL METHODOLOGY AND APPLICATIONS BRANCH. **Joinpoint Regression Program, Version 4.9.1.0**, April 2022.

NCI - NATIONAL CANCER INSTITUTE. **Joinpoint Trend Analysis Software National Cancer Institute Overview and Mission**, 2022. Disponível em: < <https://surveillance.cancer.gov/joinpoint/>>. Acesso em: 04 ago. 2022.

NCI - NATIONAL CANCER INSTITUTE. **National Cancer Institute Overview and Mission**, 2018. Disponível em: < <https://www.cancer.gov/about-nci/overview> >. Acesso em: 04 ago. 2022.

NCI - NATIONAL CANCER INSTITUTE. **Number of joinpoints**. Disponível em: <<https://surveillance.cancer.gov/help/joinpoint/setting-parameters/method-and-parameters-tab/number-of-joinpoints>>. Acesso em: 10 ago. 2022c.

NG, R. T.; HAN, J. CLARANS: a method for clustering objects for spatial data mining. **IEEE Transactions on Knowledge and Data Engineering**, v. 14, n. 5, p. 1003–1016, 2002.

OGP – OPEN GOVERNMENT PARTNERSHIP. Disponível em: <<https://www.opengovpartnership.org/>>. Acesso em 03 abr. 2023.

OKAMOTO, M. et al. Social capital and suicide: An ecological study in Tokyo, Japan. **Environmental Health and Preventive Medicine**, v. 18, n. 4, p. 306–312, 2013.

ORELLANA, J. D. et al. Spatial-temporal trends and risk of suicide in Central Brazil: An ecological study contrasting indigenous and non-indigenous populations. **Revista Brasileira de Psiquiatria**, v. 38, n. 3, p. 222–230, 2016.

PALMA, D. C. A.; SANTOS, E. S.; IGNOTTI, E. Analysis of spatial patterns and characterization of suicides in Brazil from 1990 to 2015. **Cadernos de Saude Publica**, v. 36, n. 4, p. 1–13, 2020.

PHA – PLANETARY HEALTH ALLIANCE. Disponível em: < <https://www.planetaryhealthalliance.org/>>. Acesso em 10 fev. 2023.

PROGRAMA DAS NAÇÕES UNIDAS PARA O DESENVOLVIMENTO – BRASIL. **O que é o IDHM**, c2020. Disponível em: < <https://www.br.undp.org/content/brazil/pt/home/idh0/conceitos/o-que-e-o-idhm.html> >. Acesso em: 10 mai. 2020.

PUBMED. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/>. Acesso em: 04 ago. 2022.

QI, X.; TONG, S.; HU, W. Preliminary spatiotemporal analysis of the association between socio-environmental factors and suicide. **Environmental health : a global access science source**, v. 8, p. 46, 2009.

RODRIGUES, C. D. et al. Trends in suicide rates in Brazil from 1997 to 2015. **Brazilian Journal of Psychiatry**, 2019.

SALDIVA, P. H. N. Currículo do sistema currículo Lattes. [São Paulo], 22 mai. 2023. Disponível em: <<http://lattes.cnpq.br/5450805572417395>>. Acesso em: 21 out. 2023.

SAMAN, D. M. et al. Does place of residence affect risk of suicide? a spatial epidemiologic investigation in Kentucky from 1999 to 2008. **BMC Public Health**, v. 12, n. 1, p. 108, 2012.

SANTOS, E. G. O.; BARBOSA, I. R.; SEVERO, A. K. S. Space-time analysis of mortality by suicide in the state of Rio Grande do norte, Brazil, in the period from 2000 to 2015. **Ciencia e Saude Coletiva**, v. 25, n. 2, p. 633–643, 2020.

SCANAVINO, M d T. **Currículo do sistema currículo Lattes**. [São Paulo], 31 ago. 2023. Disponível em: <<http://lattes.cnpq.br/0164156685942238>>. Acesso em: 21 out. 2023.

SMITH, S. M. et al. Correspondence of the brain's functional architecture during activation and rest. NATIONAL ACADEMY OF SCIENCES. **Proceedings** [...]. United States of America, 2009, v. 106, n. 31, p. 13040–13045.

SUICITIES. Disponível em: <<https://github.com/sheiladada/suicities>>. Acesso em 08 ago. 2022.

TABLEAU. Tableau Public, 2022. Disponível em <https://public.tableau.com/s/>. Acesso em 08 ago. 2022.

TAKAHASHI, K. et al. A flexibly shaped space-time scan statistic for disease outbreak detection and monitoring. **International Journal of Health Geographics**, v. 7, p. 1–14, 2008.

TABNET. Mortalidade geral – 1996 a 2015: Notas Técnicas, 2017. Disponível em: <http://tabnet.datasus.gov.br/cgi/sih/Morb_geral_loc_res_2008.pdf>. Acesso em: 03 mai. 2020.

TABNET. Morbidade Hospitalar do SUS por local de residência – a partir de 2008: Notas Técnicas, 2008a Disponível em: <http://tabnet.datasus.gov.br/cgi/sih/Morb_geral_loc_res_2008.pdf>. Acesso em: 25 abr. 2020.

TABNET. Morbidade Hospitalar do SUS por local de residência – Base de Dados, 2008b. Disponível em: <<http://tabnet.datasus.gov.br/cgi/defthtm.exe?sih/cnv/nrbr.def>>. Acesso em: 25 abr. 2020.

THE WORLD BANK. Suicide mortality rate (per 100,000 population) – Brazil. Disponível em: <<https://data.worldbank.org/indicator/SH.STA.SUIC.P5?end=2019&locations=BR&start=2000&view=chart>>. Acesso em: 23 set. 2023.

TRAN, F.; MORRISON, C. Income inequality and suicide in the United States: A spatial analysis of 1684 U.S. counties using geographically weighted regression. **Spatial and Spatio-temporal Epidemiology**, v. 34, p. 100359, 2020.

TRICCO, A. C. et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. **Ann Intern Med**, 169(7):467-473, 2018. doi:10.7326/M18-0850.

TRIPLEPI, G. et al. Stratification for confounding-part 2: Direct and indirect standardization. **Nephron - Clinical Practice**, v. 116, n. 4, p. 2–5, 2010.

VAN ORDEN, K. A. et al. The Interpersonal Theory of Suicide. **Psychological Review**, v. 117, n. 2, p. 575–600, 2010.

WEB OF SCIENCE. Disponível em: <https://www.webofscience.com/>. Acesso em: 04 ago. 2022.

WHITMEE, S. et al. Safeguarding human health in the Anthropocene epoch: Report of the Rockefeller Foundation-Lancet Commission on planetary health. **The Lancet**, v. 386, n. 10007, p. 1973–2028, 2015.

WHO - WORLD HEALTH ORGANIZATION. **Suicide prevention**, c2020a. Disponível em: <<https://www.who.int/health-topics/suicide>>. Acesso em: 11 abr. 2020.

WHO - WORLD HEALTH ORGANIZATION. **Age-standardized suicide rates (per 100 000 population), both sexes**, 2016. Disponível em: <https://www.who.int/gho/mental_health/suicide_rates>. Acesso em: 11 abr. 2020.

WHO - WORLD HEALTH ORGANIZATION. **Classification of Diseases (ICD)**, c2020b. Disponível em: <<https://www.who.int/classifications/icd/en/>>. Acesso em: 27 abr. 2020.

WHO - WORLD HEALTH ORGANIZATION. Comprehensive Mental Health Action Plan 2013-2030. **Report**. Geneva, 2021.

WHO - WORLD HEALTH ORGANIZATION. **Facts and figures about suicide** Geneva: World Health Organisation Press Office, 1999. Disponível em: <<http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Facts+and+Figures+About+Suicide#2>>

WHO - WORLD HEALTH ORGANIZATION. Methods and data sources for country-level causes of death 2000-2016. **World Health Organization**, December, p. 1, 2018.

WHO - WHO WORLD HEALTH ORGANIZATION. **Age-standardized suicide rates (per 100 000 population)**. Disponível em: <[https://www.who.int/data/gho/data/indicators/indicator-details/GHO/age-standardized-suicide-rates-\(per-100-000-population\)](https://www.who.int/data/gho/data/indicators/indicator-details/GHO/age-standardized-suicide-rates-(per-100-000-population))>. Acesso em: 24 jun. 2021.

WILKINSON, M. D. et al. Comment: The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, v. 3, p. 1–9, 2016.

YAMAOKA, K. et al. Spatial clustering of suicide mortality and associated community characteristics in Kanagawa prefecture, Japan, 2011-2017. **BMC Psychiatry**, v. 20, n. 1, p. 1–15, 2020.

YEH, C.-C. M. et al. Matrix Profile I: All Pairs Similarity Joins for Time Series: A Unifying View That Includes Motifs, Discords and Shapelets. **IEEE International Conference on Data Mining (ICDM)**. December, p. 1317–1322, 2017.

YU, J. et al. Seasonality of suicide: a multi-country multi-community observational study. **Epidemiology and psychiatric sciences**, v. 29, p. e163, 2020.

ZHAO, L. et al. Multi-task learning for spatio-temporal event forecasting. In: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, **Proceedings** [...]. Sydney NSW Australia, 2015, v. 2015- August, p. 1503–1512.

ZHOU, Z.; MATTESON, D. S. Predicting ambulance demand: A spatio-temporal kernel approach. In: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, **Proceedings** [...]. Sydney NSW Australia, 2015, v. 2015- August, p. 2297–2303.

ZHU, Y. et al. Matrix profile II: Exploiting a novel algorithm and GPUs to break the one hundred million barrier for time series motifs and joins. In: IEEE INTERNATIONAL CONFERENCE ON DATA MINING, ICDM. **Proceedings** [...]. Barcelona, Spain, 2017, p. 739–748.

GLOSSÁRIO

Agrupamento (clustering) – Técnica de aprendizado estatístico ou de máquina não supervisionado. Ou seja, não há variável resposta para treinar o algoritmo, as entidades se agrupam de acordo com a semelhança entre suas variáveis.

APC (Annual Percent Change - Taxa de mudança anual) – Percentual de aumento ou queda do suicídio de uma série histórica.

CID – Classificação Internacional de Doenças.

DATASUS – Departamento de Informática do Sistema Único de Saúde.

Elbow – Ponto de inflexão no gráfico que apresenta o número de agrupamentos formados e a soma da distância quadrada de cada ponto até seu centroide para cada número de agrupamentos.

FAIR – princípios de gestão de dados para aumentar o reuso de bases de dados acadêmicas. Os dados devem ser encontráveis, acessíveis, interoperáveis e reutilizáveis (*Findable, Accessible, Interoperable, and Reusable*).

Joinpoint – Número de “quebras”, “cortes”, regressões que são feitas para modelar uma série histórica em uma determinada quantidade de regressões.

NCI - National cancer institute - Instituto Nacional do Câncer dos Estados Unidos

SIM – Sistema de Informação sobre Mortalidade.

Taxa padronizada por idade – Taxa da incidência de um evento caso apresentasse uma estrutura etária padrão.