

BrSuicides: Conjunto de dados para análise de casos diários de suicídio no Brasil

Victor William Klann¹, Elaine Parros Machado de Sousa¹

¹Universidade de São Paulo
Instituto de Ciências Matemáticas e de Computação
Av. Trab. São Carlsense, 400 - Centro, São Carlos - SP, 13566-590

victor.klann@usp.br, parros@icmc.usp.br

Abstract. *Suicide is among the main causes of death in the world and with increasing rates in Brazil. The WHO and health professionals argue that the collection and analysis of suicide data are important for implementing and evaluating regulations and public policies to assist in suicide prevention. Brazil, however, still lacks in-depth analysis and implementation of public policies for suicide prevention. In this context, this work presents and makes available the BrSuicides set, composed of data related to suicide cases in Brazil and created from the death database provided by DATASUS. The pre-processing tasks significantly reduced the inconsistencies presented in the data found and examined in a set prepared for analysis from different perspectives, from exploratory statistical analysis to time series machine learning.*

Resumo. *O suicídio está entre as principais causas de morte no mundo e com taxas crescentes no Brasil. A OMS e profissionais de saúde defendem que a coleta e a análise dos dados de suicídio são importantes para implementação e avaliação de regulações e políticas públicas para auxiliar na prevenção do suicídio. O Brasil, no entanto, ainda carece de análises aprofundadas e de implementação de políticas públicas para prevenção de suicídio. Nesse contexto, este trabalho apresenta e disponibiliza o conjunto BrSuicides, composto por dados relacionados a casos de suicídio no Brasil e criado a partir da base de dados de óbitos fornecida pelo DATASUS. As tarefas de pré-processamento reduziram significativamente as inconsistências presentes nos dados coletados e resultou num conjunto preparado para análise sob diferentes perspectivas, de análise estatística exploratória a aprendizado de máquina em séries temporais.*

1. Introdução

O suicídio tem se tornado um problema sério de saúde pública global e é estimado que mais de 700.000 pessoas cometem suicídio todos os anos (WHO 2021). O suicídio é definido como um ato de lesão ou intoxicação autodirigida intencionalmente (WHO 1992). No Brasil, foi observado um aumento crescente nos casos de suicídio desde os anos 2000 (Ilic and Ilic 2022), e há uma falta de programas governamentais que trabalhem efetivamente na prevenção do suicídio (da Mata et al. 2020). O estudo desse fenômeno complexo é necessário para o desenvolvimento de políticas públicas que promovam o enfrentamento desse problema e de sua prevenção (SVS 2021). Portanto, a análise de dados de suicídio possibilita a avaliação dos processos de monitoramento e prevenção atuais baseada em evidências e dados confiáveis (WHO 2021; SVS 2021).

O DATASUS¹ disponibiliza dados provenientes de diferentes sistemas do Ministério da Saúde (MS), como o Sistema de Informação sobre Mortalidade (SIM) e o Sistema de Informação de Agravos de Notificação (SINAN). Soluções como as apresentadas em (Coelho et al. 2021) e (Petruzalek 2016), desenvolvidas para as linguagens Python e R respectivamente, buscam facilitar o acesso aos dados nos repositórios do DATASUS, disponibilizando alguns critérios de seleção, como anos e unidades federativas de interesse. Em (Gonçalves et al. 2021), por exemplo, os autores publicaram uma base de dados relacionada à campanha de vacinação do COVID-19 no Brasil, criada a partir de dados disponibilizados pelo DATASUS. Os autores também destacam os desafios na utilização desses dados, indicando alto volume de dados faltantes e dados com preenchimento errôneo.

Nesse cenário, este trabalho apresenta o BrSuicides, um conjunto de dados referentes a casos de suicídio no Brasil, ocorridos no período de 1996 a 2022. O conjunto foi criado a partir de dados de casos de óbito coletados do DATASUS, pré-processados para redução de inconsistências e organizados com o objetivo de apoiar análises tanto estatísticas quanto baseadas em aprendizado de máquina, com ou sem informação temporal. O conjunto permite a criação de séries temporais com diferentes granularidades temporais e espaciais, por meio da agregação do número de casos de óbitos por município, estado, região ou país, por dia, mês ou ano. Essas diferentes perspectivas viabilizam a utilização de abordagens diversas, de acordo com o estudo de interesse.

Os resultados iniciais apresentados neste trabalho ilustram o potencial de uso do conjunto BrSuicides em tarefas de análise exploratória baseada em séries temporais. Vale mencionar que durante o período de desenvolvimento desta pesquisa, não foram encontradas referências de criação e compartilhamento de bases de dados pré-processadas para casos de óbito por suicídio no Brasil.

As seguintes seções apresentam os passos realizados para a extração dos dados de casos de suicídio, as etapas de pré-processamento aplicadas, o conjunto de dados resultante e exemplos de uso dos dados de forma temporal para aplicação de técnicas de análise de séries temporais.

2. Metodologia

Os dados coletados no DATASUS para criação do conjunto BrSuicides são dos casos de óbito no Brasil ocorridos entre 1996 e 2022, que é o último ano disponível de dados consolidados, verificado manualmente no servidor FTP do DATASUS¹. A decisão de coletar os dados de óbito a partir de 1996 se dá pela utilização da 10ª Revisão da Classificação Internacional de Doenças (CID-10), sendo que até 1995 a CID-9 era utilizada². Essa mudança gera um problema de compatibilidade entre as duas classificações, por terem mudanças consideráveis (WHO 1992) e o DATASUS não disponibiliza uma lista integrada de conversão. A granularidade dos dados é de natureza diária, baseada na declaração de óbito (DO) individual registrada em um Cartório de Registro Civil. Os registros dos cartórios são recolhidos pelas Secretarias de Saúde estaduais e agregadas pelo SIM. A consolidação da base de dados nacional é realizada somente após a agregação dos dados de casos

¹<ftp.datasus.gov.br/dissemin/publicos/SIM/CID10/> acesso em: 11 jan. 2024

²http://tabnet.datasus.gov.br/cgi/sim/Mortalidade_Geral_1996_2012.pdf
Acesso em: 3 mar. 2024

de óbito de todas as unidades federativas².

O servidor do DATASUS disponibiliza arquivos de extensão .DBC, versão compactada de arquivo .DBF. Cada arquivo contém os dados de um ano de um estado brasileiro, totalizando 729 arquivos para o período em questão¹. Um problema dessa extensão é a falta de compatibilidade com ferramentas de extração de dados por ser um formato antigo e possuir variações estruturais no padrão dependendo da versão usada. Para resolver esse problema de leitura, foi utilizada a biblioteca de Python *pySUS* (Coelho et al. 2021), que realiza a conversão de arquivos .DBC para .DBF e os disponibiliza para *download*. Para a obtenção e manipulação dos dados, foi utilizada a linguagem de programação Python e as bibliotecas *Pandas*³ e *NumPy*⁴.

2.1. Extração dos casos de suicídio

Para selecionar somente os casos de óbito por suicídio, os códigos CID-10 (10^a Revisão de Classificação Internacional de Doenças) referentes a esse tipo de óbito foram utilizados. O código CID-10 é composto por uma letra e dois dígitos, representando a causa principal do tipo de mortalidade. Pode-se incluir até 4 novos dígitos para representar detalhes como subcategoria, severidade, local de ocorrência e outros subtipos, ficando a critério dos órgãos de saúde optarem por utilizá-los ou não (WHO 1992). Os códigos de suicídio começam com a letra X, referente ao capítulo XX do CID-10 sobre causa externas de morbidade e mortalidade, e os dois primeiros dígitos devem estar entre 60 e 84 (WHO 1992). O terceiro dígito é uma subclassificação realizada pelo DATASUS para indicar o local de ocorrência do suicídio⁵. Por exemplo, o código X700 significa que, de acordo com os dois primeiros dígitos, 70, o suicídio foi cometido por uma lesão auto-provocada intencionalmente por enforcamento, estrangulamento e sufocação, e de acordo com o terceiro dígito, 0, ocorreu na residência do falecido.

O *download* dos dados, utilizando o *pySUS*, foi realizado selecionando todas as UFs do Brasil, escolhendo o intervalo de 1996 até 2022, e resultou em 268.028 registros com 103 atributos, mostrando uma disparidade em relação ao número de atributos possíveis de acordo com documentos oficiais do Ministério da Saúde e da Secretaria de Vigilância em Saúde e Ambiente (SVSA)⁶, que indicam a existência de 132 atributos. Ao avaliar, na documentação, a natureza de cada atributo, notou-se que 83 dos 132 atributos possíveis são irrelevantes para análises relacionadas ao tema suicídio, por serem considerados atributos administrativos ou relevantes somente para casos de óbito não relacionados a suicídio. Desses 83 atributos, 69 estavam presentes nos registros adquiridos e, após sua remoção, restaram 34 atributos.

2.2. Pré-processamento dos dados

Os registros presentes no DATASUS apresentam diversas inconsistências, referentes a fatores como: número diferente de atributos de acordo com o estado e ano, quantidade expressiva de valores nulos na maioria dos atributos, valores nominais e ordinais que não

³<https://pandas.pydata.org/>

⁴<https://numpy.org/>

⁵<http://www2.datasus.gov.br/cid10/V2008/cid10.htm> acesso em: 6 jul. 2024

⁶https://diaad.s3.sa-east-1.amazonaws.com/sim/Mortalidade_Geral++-Estrutura.pdf

estão incluídos no dicionário de conversão do seu atributo, diferentes atributos que representam a mesma informação e dados mal formatados. A estrutura geral das declarações de óbito (DO) são divulgadas pelo SIM⁶.

A limpeza e transformação dos dados foi realizada com o objetivo de reduzir tais inconsistências. Primeiramente, dos campos que armazenam texto nos registros, foram removidos espaços em branco que mascaram valores nulos, para detectá-los com maior facilidade. Em seguida, o atributo "DUTOBITO", responsável por armazenar a data do óbito, foi avaliado para remoção de registros com datas inválidas, que não poderiam ser utilizados em análises temporais (foram removidos 52 registros).

Os demais atributos foram analisados para obter a porcentagem de valores nulos de cada um, sendo que 11 deles ("IDADEMAE", "ESMAE", "OCUPMAE", "QTDFILVIVO", "QTDFILMORT", "GRAVIDEZ", "GESTACAO", "PARTO", "PESO", "ESMAE2010" e "ESMAEAGR1") possuem mais de 99% de valores nulos e 4 ("CODBAIRES", "UFINFORM", "CODBAIOCOR" e "CODMUNCART") ultrapassam o limite arbitrário de 50% de valores faltantes desde a sua data de adição, e portanto foram removidos. Dois atributos, "CODMUNNATU", que armazena o código do município de naturalidade do falecido, e "HORAOBITO", que possui a hora de ocorrência do óbito, foram removidos do conjunto de dados pela porcentagem de nulos presentes nos registros desde o início do período de coleta, em 1996, de 58% e 44,4%, respectivamente, pois não existem atributos compatíveis com a informação desses atributos que poderiam ser usados para substituir os valores faltantes.

O atributo "SEXO" possui três valores em seu dicionário: "masculino", "feminino" e "Ignorado", com "Ignorado" equivalendo a casos especiais como cadáveres mutilados, em estado avançado de decomposição, genitália indefinida ou hermafroditismo⁵. Os 55 registros de "Ignorado" não foram removidos. Os atributos "RACACOR" e "ESTCIV" também compartilham o mesmo contexto de "SEXO", possuem valores nominais com um dicionário de conversão para dados categóricos com "Ignorado" representando seus valores nulos. Os dois atributos contavam com 21.535 e 8.367 valores faltantes, convertidos para "Ignorado", ficando com 21.536 e 20.080 registros deste valor.

Para a informação sobre escolaridade, existem os atributos "ESC", "ESC2010" e "ESCFALAGR1", que armazenam o nível de escolaridade do falecido, incluídos na DO em 1996, 2011 e 2012, respectivamente. Apesar de "ESC2010" e "ESCFALAGR1" apresentarem alta porcentagem de valores nulos (casos de óbitos anteriores a 2011), o atributo "ESC" pode ser usado para determinar a escolaridade até a data de implementação de "ESC2010" e "ESCFALAGR1", e esses são utilizados para os registros mais recentes. Para padronizar os valores antigos e recentes dos atributos, foi aplicada uma conversão dos dados para se adequarem com os intervalos de anos especificados no atributo "ESC", com base na tabela de conversão divulgada pela Coordenação Geral de Informações e Análise Epidemiológica (CGIAE 2011). Por exemplo, em "ESCFALAGR1", o valor "04" representa "Fundamental II completo", "ESC2010" possui o valor "3" que representa "Médio (antigo 2º Grau)". Para padronizar esses valores, o valor "8 a 11 anos" foi usado, pois completar o fundamental II requer 8 anos de estudo, e o ensino médio acontece entre 9 e 11 anos de estudo. Para o valor "07" de "ESCFALAGR1", que representa "Superior Incompleto", o valor usado é "12 anos e mais", pois inclui os anos acumulados até no mínimo um ano de ensino superior, resultando em pelo menos 12 anos de estudo. No

atributo "ESC" havia um valor redundante devido a mudanças em formulários de DO², "9 a 11 anos", que foi convertido para "8 a 11 anos".

Para tratar os dados referentes à idade da vítima, o atributo "IDADE" foi utilizado. Os valores presentes foram convertidos conforme a descrição presente no documento de estrutura⁵. Estes valores são compostos por um subcampo de um dígito, indicando se a idade é definida em minutos, horas, meses ou anos, seguido de um segundo subcampo que indica a quantidade de unidades do tempo. Por exemplo, nos valores "433" e "512" o dígito 4 manifesta que a idade está em anos, e 5 indica que a idade está em anos e é maior do que 100 anos. Logo, a primeira idade converte-se para 33 anos e a segunda converte-se para $100 + 12 = 112$ anos. Para os valores inválidos, utilizou-se a diferença entre a data do óbito em "DTOBITO" e a de nascimento em "DTNASC" para extrair uma idade válida ou faltante. Registros com idades abaixo de 5 anos foram removidos, conforme realizado em (SVS 2021), assim como os registros faltantes que não possibilitaram extrair uma idade válida, uma vez que "IDADE" e "DTNASC" tinham valores inválidos ou faltantes, sendo removidos 768 registros.

O atributo "OCUP", responsável por armazenar a ocupação do falecido, de acordo com a Classificação Brasileira de Ocupações (CBO), apresenta códigos inválidos em 66.724 registros, e valores faltantes em 49.476, somando 116.200 registros. Como a quantidade de valores nulos e inválidos é significativa, optou-se por manter tanto os valores nulos, convertidos para "0", quanto os inválidos, para evitar a perda de observações. Os códigos válidos foram convertidos para as respectivas ocupações a partir da CBO.

Também foram tratados os atributos "CODMUNOCOR" e "CODMUNRES", que mantêm os códigos de município de residência do falecido e de ocorrência do óbito, respectivamente, baseado em dois arquivos que possuem códigos de municípios, sendo o primeiro disponibilizado pelo Sistema Único de Assistência Social (SUAS)⁷ e o segundo disponibilizado pela Divisão Territorial Brasileira (DTB)⁸, convertido para o formato .csv (*comma-separated values*). Ambos os arquivos foram utilizados para converter o código dos municípios para seus respectivos nomes. Após a conversão, foram detectados códigos inválidos em 468 registros em "CODMUNOCOR", e após sua remoção, foram detectados 513 códigos inválidos em CODMUNRES, que também foram removidos. Essa limpeza resultou em 981 registros retirados.

O atributo "NATURAL" detém o código do país e da UF onde o falecido nasceu e caso este seja estrangeiro, o código do país é informado. Para casos no Brasil, o código inicia pelo número 8, seguido do código da UF. Seus valores estavam armazenados como texto, sendo convertidos para números e possui 44.481 valores faltantes, com a decisão de mantê-los para evitar a perda de muitos registros. Por fim, o atributo "LOCOCOR" foi tratado. "LOCOCOR" possui o local de ocorrência do falecimento, em texto. Seu valor considerado nulo é "ignorado", pois significa que não foi possível identificar onde ocorreu o óbito. Os valores faltantes, convertidos para "ignorado", somam em conjunto com os valores existentes de "ignorado", um volume de 2.502 registros, que foram mantidos.

⁷<http://blog.mds.gov.br/redesuas/lista-de-municipios-brasileiros/> acesso em: 08 jul. 2024

⁸<https://www.ibge.gov.br/geociencias/organizacao-do-territorio/estrutura-territorial/23701-divisao-territorial-brasileira.html> acesso em: 10 jul 2024

2.3. Conjunto de dados resultante

Após o pré-processamento dos dados coletados, o conjunto gerado, nomeado BrSuicides, inclui 266.236 registros e 17 atributos. A tabela 1 expõe o volume e porcentagem de valores nulos presentes em cada coluna, enquanto que tabela 2 descreve o tipo dos dados de cada atributo, e uma breve explicação de o que eles representam. Os valores considerados nulos para atributos numéricos são *Not a Number* (NaN), e para atributos categóricos são considerados os valores explorados na metodologia, como o valor "0" que representa o valor nulo do atributo "OCUP".

A tabela 1 descreve o tipo dos dados de cada atributo, e uma breve explicação de o que eles representam. A tabela 2 mostra a porcentagem final de 5 tipos de valores de cada atributo:

1. Faltantes ou nulos: são registros vazios ou que possuem um valor que representa a falta de informação, como "9" para "SEXO" e "0" em "OCUP", representando a escolha de não preencher o registro. Os valores refletem a presença destes valores antes de qualquer manipulação do atributo;
2. Inválidos: valores inválidos não tem uma conversão que permita sua utilização, como em "CODMUNRES" e "CODMUNOCOR" que possuem códigos de municípios que não podem ser convertidos em um município. Os valores podem refletir a descoberta desses valores antes ou depois da manipulação dos dados, como uma tentativa de conversão para outro tipo de dado. Um exemplo disso é "DTNASC", em que primeiro foram detectados valores faltantes, e após a tentativa de conversão do valor armazenado para um formato YY-mm-dddd de data, foram detectadas datas inválidas;
3. Tratados: valores que foram alterados para facilitar ou viabilizar seu uso, como no caso da "IDADE" que exigia a conversão presente na sua descrição para extrair uma idade válida⁶. Também abrangem os valores faltantes, inválidos e nulos que foram alterados de alguma forma, como "SEXO", "RACACOR" e "ESTCIV" que tiveram seus valores faltantes convertidos para nulos; e
4. Removidos: dados que foram considerados desnecessários para a base de dados, como os dados faltantes, nulos e inválidos que não foram alterados ou mantidos.

Como a remoção de valores altera o número de registros presentes no conjunto de dados, as porcentagens representam o tamanho do conjunto no momento da execução do código. Os números presentes na tabela também podem diferir dos mencionados na seção anterior devido ao mesmo motivo e também pelas diferentes etapas de pré-processamento realizadas em alguns atributos, como "SEXO" e "LOCOCOR".

Tabela 1. Descrição dos atributos presentes em BrSuicides

Atributo	Descrição	Tipo do dado
ESTADO	Estado de ocorrência do óbito.	Texto
ANO	Ano de ocorrência do óbito.	Números
DTOBITO	Data em que ocorreu o óbito.	Data (yyyy-mm-dd)
NATURAL	País e Unidade da Federação onde falecido nasceu. Se estrangeiro informar País.	Números
DTNASC	Data do nascimento do falecido.	Data (yyyy-mm-dd)
IDADE	Idade do falecido em anos.	Números
SEXO	Sexo do falecido.	Texto
RACACOR	Cor informada pelo responsável pelas informações do falecido.	Texto
ESTCIV	Situação conjugal do falecido informada pelos familiares.	Texto
ESC	Escolaridade em anos.	Texto
OCUP	Tipo de trabalho que o falecido desenvolveu na maior parte de sua vida produtiva.	Texto
CODMUNRES	Município de residência do falecido.	Texto
LOCOCOR	Local de ocorrência do óbito.	Texto
CODMUNOCOR	Município onde ocorreu o óbito.	Texto
CAUSABAS	Causa básica da Declaração de Óbito. (Códigos CID 10)	Texto
ESC2010	Escolaridade 2010. Nível da última série concluída pelo falecido.	Texto
ESCFALAGRI	Escolaridade do falecido agregada (formulário a partir de 2010).	Texto

Tabela 2. Porcentagem dos tipos de dados dos atributos de BrSuicides

Atributos	% Faltantes/Nulos	% Inválidos	% Tratados	% Removidos	Tam. Dataset
DTOBITO	0.00	0.02	100.00	0.02	268028
ESTADO	0.00	0.00	100.00	0.00	267976
ANO	0.00	0.00	100.00	0.00	267976
SEXO	0.03	0.00	100.00	0.00	267976
RACACOR	15.99	0.00	100.00	0.00	267976
ESTCIV	14.76	0.00	95.85	0.00	267976
DTNASC	0.73	0.79	0.79	0.00	267976
IDADE	0.28	0.00	99.71	0.29	267208
ESC	18.72	15.25	66.03	0.00	267208
ESC2010	8.03	52.13	39.84	0.00	267208
ESCFALAGRI	6.77	58.91	34.32	0.00	267208
OCUP	18.52	24.97	75.03	0.00	267208
CODMUNOCOR	0.00	0.18	99.82	0.18	266227
CODMUNRES	0.00	0.19	99.81	0.19	266227
NATURAL	16.71	0.00	83.29	0.00	266227
LOCOCOR	0.80	0.15	99.85	0.00	266227

Os dados estão disponíveis em (BrSuicides 2024), em conjunto com os códigos necessários para reproduzir a limpeza e modelagem dos dados e as aplicações apresentadas.

2.4. Trabalhos Futuros

Para melhorar o gerador deste conjunto de dados, é possível realizar as seguintes etapas na extração dos dados e na documentação dos processos realizados:

1. Incluir no código de extração dos dados pelo pySUS para realizar o *download* de todos os anos a partir de 1996 sem necessidade de incluir manualmente anos mais recentes;
2. Construir um *pipeline* para automatizar o processo de extração e pré-processamento dos dados para gerar o conjunto de dados; e

3. Adicionar tabelas e arquivos informativos no repositório do Github que auxiliem no entendimento dos atributos, seus valores e da metodologia aplicada.

3. Análise exploratória

Com este conjunto de dados, foram aplicados diferentes tipos de análises temporais. Os dados utilizados são dados absolutos de casos de óbito por suicídio. Inicialmente, é natural que a agregação dos dados diários sejam realizados, pois as análises de suicídio costumam ser realizadas a partir da contagem de casos de suicídios em um intervalo de tempo. A conversão do conjunto de dados para um formato de séries temporais pode ser realizada de forma simples, utilizando os valores de "DTOBITO" como índices das séries e agregando os casos de suicídios que ocorreram em uma mesma data para gerar a contagem dos casos na data em questão. As figuras 1 e 2 expressam a aplicação de diferentes tipos de agregações, uma diária e outra mensal, para a visualização dos dados do Brasil durante o período coletado de 1996 a 2022.

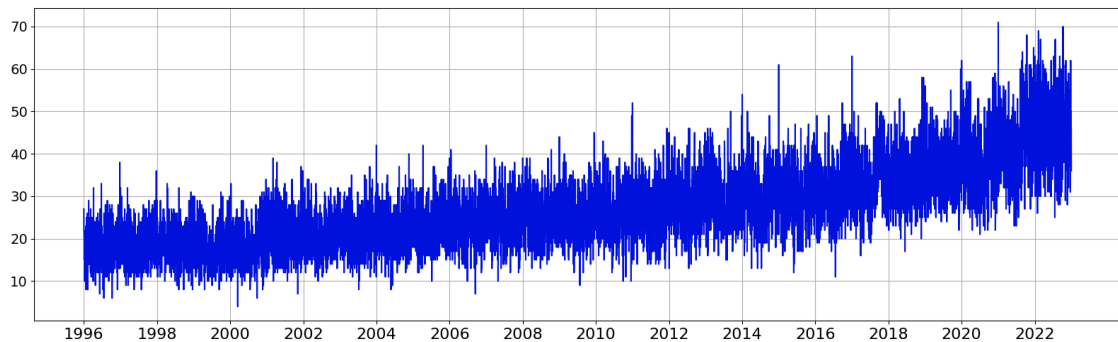


Figura 1. Frequência diária de ocorrências de casos de suicídio no Brasil entre 1996 e 2022

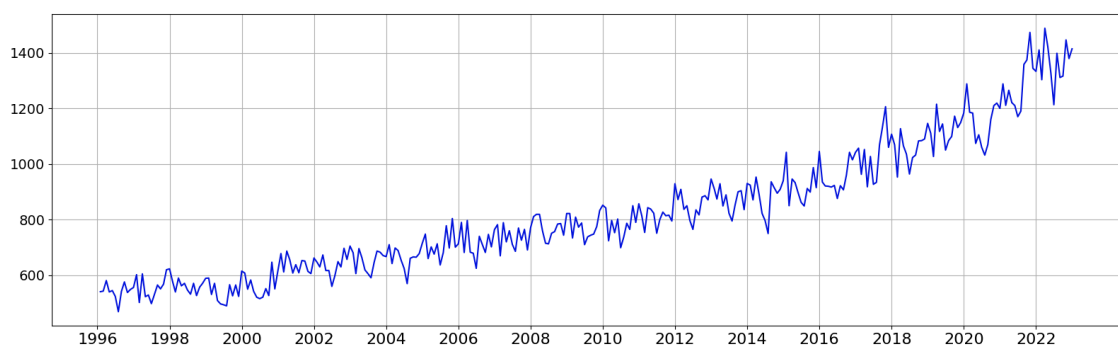


Figura 2. Frequência mensal de ocorrências de casos de suicídio no Brasil entre 1996 e 2022

Foram também analisados atributos específicos e as diferenças de comportamentos entre seus valores, assim como a filtragem dos dados de estados e regiões, como mostra a figura 3, para casos de suicídio no nordeste separados por sexo. Os casos de suicídio cometidos por ambos os sexos continuaram crescentes desde 1996, e que a taxa de suicídio pelo sexo masculino é consideravelmente maior que a pelo sexo feminino,

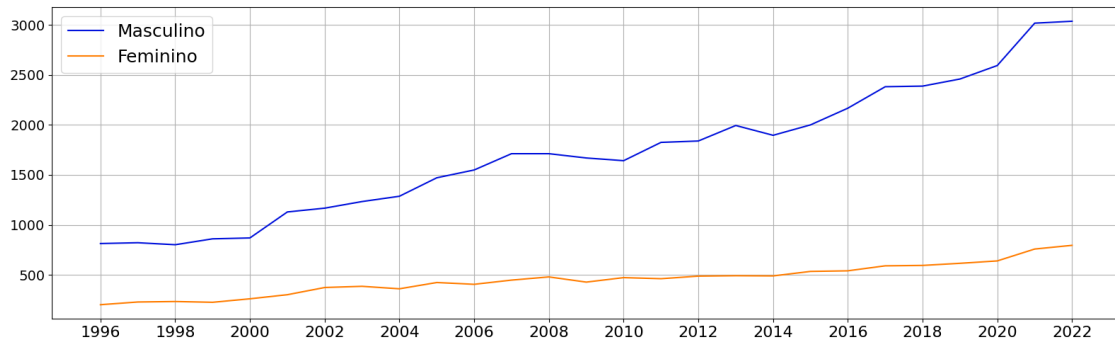


Figura 3. Casos anuais de suicídio no nordeste por sexo entre 1996 e 2022

e mostra um crescimento expressivo durante os últimos anos, e apesar da tendência de casos de suicídio pelo sexo feminino ser menor, também é positiva.

Para avaliar a evolução dos dados temporais ao longo do tempo, é geralmente necessário extrair certos componentes de uma série temporal, como a sua tendência e sazonalidade. A figura 4 mostra a extração da tendência dos casos de suicídio no sudeste utilizando um modelo linear de regressão, a regressão de Poisson, e a figura 5 expõe o componente de sazonalidade da região, extraído a partir do algoritmo STL (Cleveland et al. 1990), após a aplicação de z-normalização na série temporal.

Verifica-se que a tendência manifestada dos casos de suicídio no nordeste se mantém crescente ao longo de todo o período observado, e que o comportamento da sazonalidade das observações mudou ao longo dos anos. A diferença de amplitude da variância dos dados também mostra que antigamente os dados definiam um modelo aditivo de composição da sazonalidade, que foi alterado o suficiente para representar ao menos um modelo híbrido aditivo/multiplicativo.

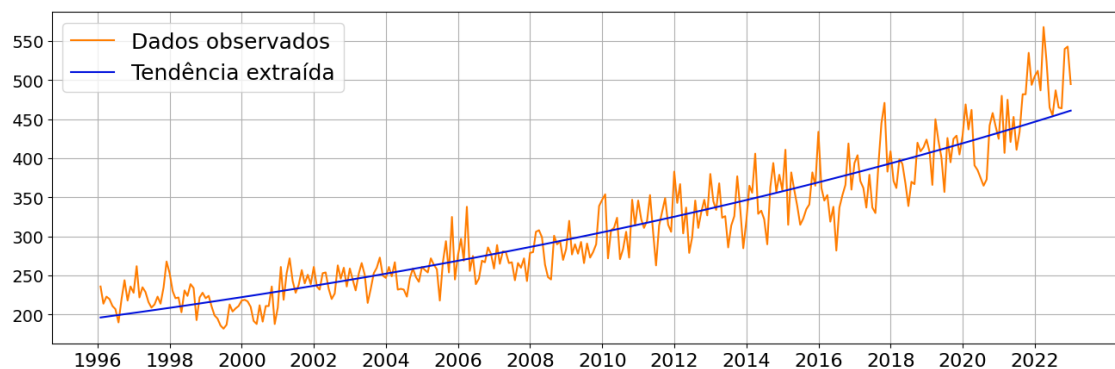


Figura 4. Tendência dos casos de suicídio do Sudeste de 1996 a 2022 extraída usando uma regressão simples de Poisson

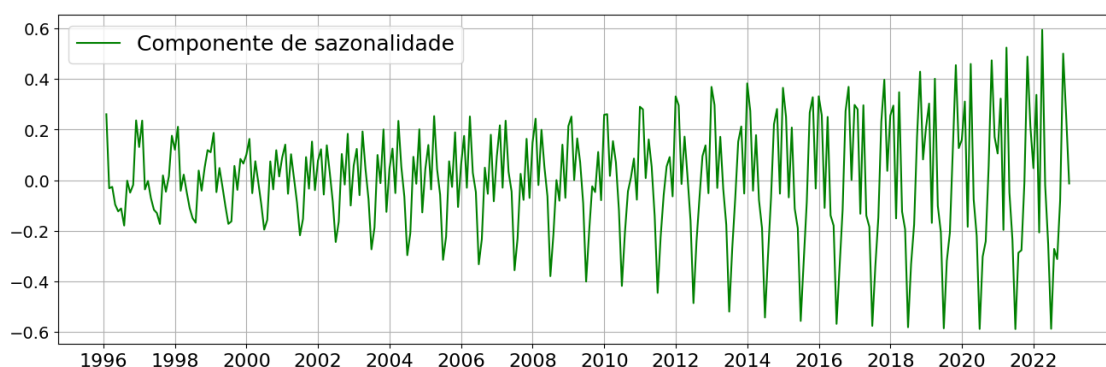


Figura 5. Sazonalidade dos casos de suicídio no sudeste entre 1996 e 2022

Outro tipo de análise que pode ser feita no contexto de séries temporais é a similaridade dos casos de suicídio entre estados e regiões, podendo ser avaliada largamente através das distâncias entre os valores de suas observações. Um ponto interessante a ser explorado é a similaridade entre um estado e sua própria região, conforme a figura 6, que exemplifica a similaridade entre o estado do Paraná e sua região, Sul, utilizando uma medida *lock-step* de comparação ponto a ponto entre as duas séries temporais, e exibe uma certa semelhança no seu comportamento de evolução na ocorrência de suicídios.

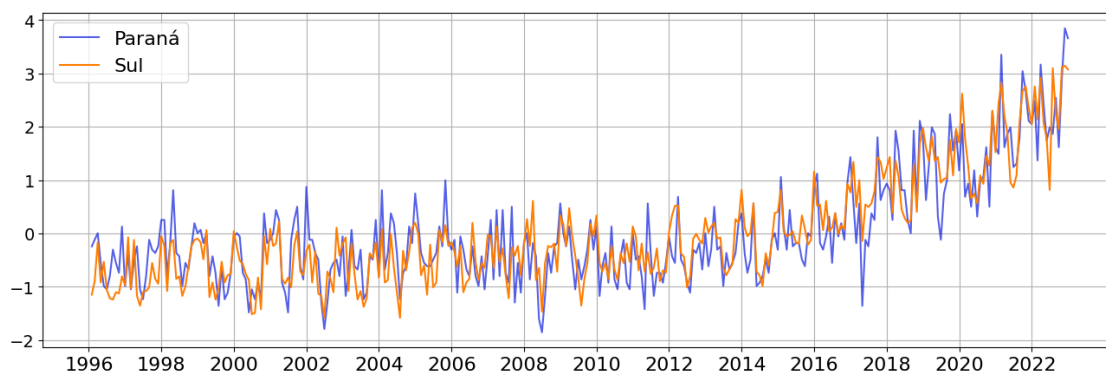


Figura 6. Avaliação da distância entre Paraná e Sul usando distância euclidiana

4. Considerações finais

Neste artigo está contido a coleta, modelagem, pré-processamento e publicação de um conjunto de dados, BrSuicides, a partir dos dados de casos de óbito causados por suicídio disponibilizados pelo DATASUS. É apresentada também uma breve análise exploratória temporal das observações de suicídio para relevar o potencial deste conjunto de dados no auxílio da exploração do suicídio no Brasil, um tema complexo que já se mostrou possível de ser enfrentado através da aplicação de boas medidas de monitoramento e prevenção, e que sua consequência direta no aumento da mortalidade pode ser reduzido, mas que no Brasil ainda carece de análises aprofundadas para auxiliar neste processo.

4.1. Agradecimentos

Este trabalho foi realizado com apoio financeiro da CAPES.

Referências

- [BrSuicides 2024] BrSuicides (2024). Brsuicides. Disponível em: <<https://github.com/vklannusp/BrSuicides-dataset>>. acesso em: 22 jul. 2024.
- [CGIAE 2011] CGIAE (2011). Consolidação da base de dados de 2011. CGIAE. Disponível em: http://tabnet.datasus.gov.br/cgi/sim/Consolidada_Sim_2011.pdf. Acesso em: 10 jun. 2023.
- [Cleveland et al. 1990] Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I. (1990). Stl: A seasonal-trend decomposition. *J. Off. Stat.*, 6(1):3–73.
- [Coelho et al. 2021] Coelho, F. C., Baron, B. C., de Castro Fonseca, G. M., Reck, P., and Palumbo, D. (2021). Alertadengue/pysus: Vaccine. Disponível em: <<https://doi.org/10.5281/zenodo.4883502>>. acesso em: 07 mai. 2023.
- [da Mata et al. 2020] da Mata, K. C. R., Daltro, M. R., and Ponde, M. P. (2020). Perfil epidemiológico de mortalidade por suicídio no brasil entre 2006 e 2015. *Revista Psicologia, Diversidade e Saúde*, 9(1):74–87.
- [Gonçalves et al. 2021] Gonçalves, M. V. F., dos Santos, J. S., Ferreira, C. Z., Zavaleta, J., da Cruz, S. M. S., and Sampaio, J. O. (2021). Datasets curados e enriquecidos com proveniência da campanha nacional de vacinação contra covid-19. In *Anais do III Dataset Showcase Workshop*, pages 148–159. SBC.
- [Ilic and Ilic 2022] Ilic, M. and Ilic, I. (2022). Worldwide suicide mortality trends (2000-2019): A joinpoint regression analysis. *World journal of psychiatry*, 12(8):1044.
- [Petruzalek 2016] Petruzalek, D. (2016). Read.dbc: um pacote para importação de dados do datasus na linguagem r. In *Anais do XV Congresso Brasileiro de Informática em Saúde [Internet]*, pages 27–30.
- [SVS 2021] SVS (2021). Boletim epidemiológico. secretaria de vigilância em saúde. *Boletim Epidemiológico*, 52:10. Disponível em: <https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/epidemiologicos/edicoes/2021/boletim_epidemiologico_svs_33_final.pdf>. acesso em: 11 mai. 2023.
- [WHO 1992] WHO (1992). The icd-10 classification of mental and behavioural disorders : clinical descriptions and diagnostic guidelines. World Health Organization. Disponível em: <https://apps.who.int/iris/handle/10665/37958>. Acesso em: 10 mai. 2023.
- [WHO 2021] WHO (2021). Suicide worldwide in 2019: global health estimates. Geneva: World Health Organization. License: CC BY-NC-SA 3.0 IGO, ISBN 978-92-4-002664-3.