

**INSTITUTO FEDERAL GOIANO – CAMPUS CERES  
BACHARELADO EM SISTEMAS DE INFORMAÇÃO  
GILSIMAR ANTÔNIO DE OLIVEIRA**

**DESCOBERTA DE CONHECIMENTO SOBRE O SUICÍDIO NO BRASIL EM UMA  
BASE DE DADOS EXTRAÍDA DO KAGGLE**

**CERES – GO  
2022**

**GILSIMAR ANTÔNIO DE OLIVEIRA**

**DESCOBERTA DE CONHECIMENTO SOBRE O SUICÍDIO NO BRASIL EM UMA  
BASE DE DADOS EXTRAÍDA DO KAGGLE**

Trabalho de curso apresentado ao curso de Sistemas de Informação do Instituto Federal Goiano – Campus Ceres, como requisito parcial para a obtenção do título de Bacharel em Sistemas de Informação, sob orientação do Prof. Dr. Rafael Divino Ferreira Feitosa.

**CERES – GO  
2022**

Sistema desenvolvido pelo ICMC/USP  
Dados Internacionais de Catalogação na Publicação (CIP)  
**Sistema Integrado de Bibliotecas - Instituto Federal Goiano**

048d Oliveira, Gilsimar Antônio de  
Descoberta de conhecimento sobre o suicídio no  
Brasil em uma base de dados extraída do kaggle /  
Gilsimar Antônio de Oliveira; orientador Dr. Rafael  
Divino Ferreira Feitosa. -- Ceres, 2022.  
28 p.

TCC (Graduação em Sistemas de Informação) --  
Instituto Federal Goiano, Campus Ceres, 2022.

1. Análise de dados. 2. Suicídio. 3. KDD. 4.  
Políticas Públicas. I. Feitosa, Dr. Rafael Divino  
Ferreira, orient. II. Título.

# TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR PRODUÇÕES TÉCNICO-CIENTÍFICAS NO REPOSITÓRIO INSTITUCIONAL DO IF GOIANO

Com base no disposto na Lei Federal nº 9.610, de 19 de fevereiro de 1998, AUTORIZO o Instituto Federal de Educação, Ciência e Tecnologia Goiano a disponibilizar gratuitamente o documento em formato digital no Repositório Institucional do IF Goiano (RIIF Goiano), sem ressarcimento de direitos autorais, conforme permissão assinada abaixo, para fins de leitura, download e impressão, a título de divulgação da produção técnico-científica no IF Goiano.

## IDENTIFICAÇÃO DA PRODUÇÃO TÉCNICO-CIENTÍFICA

- |  |   |
|--|---|
| <input type="checkbox"/> Tese (doutorado)                      | <input type="checkbox"/> Artigo científico              |
| <input type="checkbox"/> Dissertação (mestrado)                | <input type="checkbox"/> Capítulo de livro              |
| <input type="checkbox"/> Monografia (especialização)           | <input type="checkbox"/> Livro                          |
| <input checked="" type="checkbox"/> TCC (graduação)            | <input type="checkbox"/> Trabalho apresentado em evento |
| <input type="checkbox"/> Produto técnico e educacional - Tipo: |   |

Nome completo do autor:

Gilsimar Antônio de Oliveira

Matrícula:

2019103202030032

Título do trabalho:

Descoberta de conhecimento sobre o suicídio no Brasil em uma base de dados extraída do kaggle

## RESTRIÇÕES DE ACESSO AO DOCUMENTO

Documento confidencial:  Não  Sim, justifique:

Informe a data que poderá ser disponibilizado no RIIF Goiano: 15 / 12 / 2022

O documento está sujeito a registro de patente?  Sim  Não

O documento pode vir a ser publicado como livro?  Sim  Não

## DECLARAÇÃO DE DISTRIBUIÇÃO NÃO-EXCLUSIVA

O(a) referido(a) autor(a) declara:

- Que o documento é seu trabalho original, detém os direitos autorais da produção técnico-científica e não infringe os direitos de qualquer outra pessoa ou entidade;
- Que obteve autorização de quaisquer materiais inclusos no documento do qual não detém os direitos de autoria, para conceder ao Instituto Federal de Educação, Ciência e Tecnologia Goiano os direitos requeridos e que este material cujos direitos autorais são de terceiros, estão claramente identificados e reconhecidos no texto ou conteúdo do documento entregue;
- Que cumpriu quaisquer obrigações exigidas por contrato ou acordo, caso o documento entregue seja baseado em trabalho financiado ou apoiado por outra instituição que não o Instituto Federal de Educação, Ciência e Tecnologia Goiano.

Ceres, GO

Local

13 / 12 / 2022

Data

Assinatura do autor e/ou detentor dos direitos autorais

Ciente e de acordo:

Assinatura do(a) orientador(a)



SERVIÇO PÚBLICO FEDERAL  
MINISTÉRIO DA EDUCAÇÃO  
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA  
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA GOIANO

### ATA DE DEFESA DE TRABALHO DE CURSO

Ao(s) 1 dia(s) do mês de dezembro do ano de dois mil e vinte e dois, realizou-se a defesa de Trabalho de Curso do(a) acadêmico(a) Gilsimar Antonio de Oliveira, do Curso de Bacharelado em Sistemas de Informação, matrícula 2019103202030032, cujo título é “Descoberta de conhecimento sobre o suicídio no Brasil em uma base de dados extraída do sistema DATASUS do Ministério da Saúde”. A defesa iniciou-se às **18** horas e **15** minutos, finalizando-se às **19** horas e **30** minutos. A banca examinadora considerou o trabalho **APROVADO** com média 8,9 no trabalho escrito, média **9,4** no trabalho oral, apresentando assim média aritmética final de **9,2** pontos, estando o(a) estudante **APTO** para fins de conclusão do Trabalho de Curso.

Após atender às considerações da banca e respeitando o prazo disposto em calendário acadêmico, o(a) estudante deverá fazer a submissão da versão corrigida em formato digital (.pdf) no Repositório Institucional do IF Goiano – RIIF, acompanhado do Termo Ciência e Autorização Eletrônico (TCAE), devidamente assinado pelo autor e orientador.

Os integrantes da banca examinadora assinam a presente.

*(Assinado Eletronicamente)*  
Prof. Dr. Rafael Divino Ferreira Feitosa

*(Assinado Eletronicamente)*  
Prof<sup>a</sup>. Dra. Jaqueline Alves Ribeiro

*(Assinado Eletronicamente)*  
Prof. Me. Adriano Honorato Braga

Documento assinado eletronicamente por:

- Adriano Honorato Braga, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 12/12/2022 12:23:18.
- Jaqueline Alves Ribeiro, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 01/12/2022 21:58:57.
- Rafael Divino Ferreira Feitosa, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 01/12/2022 20:31:26.

Este documento foi emitido pelo SUAP em 01/12/2022. Para comprovar sua autenticidade, faça a leitura do QRCode ao lado ou acesse <https://suap.ifgoiano.edu.br/autenticar-documento/> e forneça os dados abaixo:

Código Verificador: 448861  
Código de Autenticação: cf6d68f249



INSTITUTO FEDERAL GOIANO

Campus Ceres

Rodovia GO-154, Km.03, Zona Rural, None, None, CERES / GO, CEP 76300-000

(62) 3307-7100

## RESUMO

A busca de conhecimento em base de dados (Knowledge Discovery in Databases – KDD) é uma prática cada vez mais comum para a análise de grande volume de dados. Para isso é necessário utilizar ferramentas que auxiliem na busca por padrões entre os dados. Atualmente é possível ter acesso a grandes bases de dados ou compiladas por pesquisadores ou disponibilizadas em sistemas governamentais. O kaggle é uma dessas plataformas que disponibiliza, entre outros recursos, uma enorme quantidade de *datasets* (base de dados) de diversas áreas. Nesta plataforma foi possível obter a base de dados com registros compilados sobre suicídio no Brasil entre os anos 2010 e 2019, obtidos do DATASUS, sistema vinculado ao Ministério da Saúde do Brasil. A análise dos dados foi feita com o uso da biblioteca Pandas da linguagem de programação Python. Os resultados qualitativos mostram que o uso de bibliotecas de análise de dados de alto desempenho é imprescindível para abstrair a complexidade dos modelos de análise e garantir a qualidade das informações geradas, permitindo ao especialista dedicar-se ao objeto de estudo. Os dados mostram que os casos de suicídio no Brasil vêm aumentando a cada ano, em 2010 a taxa de mortes por 100 mil habitantes era de 5,0 em 2019 saltou para 6,4. A região Sul é a que apresenta a maior taxa de mortes por suicídio, em 2019 era de 10,6 por 100 mil habitantes. A análise dos resultados mostra também uma forte relação entre a causa básica de suicídio com as variáveis idade e estado.

**Palavras-chave:** Análise de dados, Suicídio, KDD, Políticas Públicas.

## ABSTRACT

Knowledge Discovery in Databases (KDD) is an increasingly common practice for the analysis of large volumes of data. For this it is necessary to use tools that assist in the search for patterns among the data. Currently it is possible to access large databases either compiled by researchers or made available in government systems. Kaggle is one of these platforms that makes available, among other resources, a huge amount of datasets from several areas. On this platform it was possible to obtain the database with compiled records on suicide in Brazil between the years 2010 and 2019, obtained from DATASUS, a system linked to the Brazilian Ministry of Health. The data analysis was done using the Pandas library of the Python programming language. The qualitative results show that the use of high-performance data analysis libraries is essential to abstract the complexity of the analysis models and ensure the quality of the information generated, allowing the specialist to dedicate himself to the object of study. Data shows that suicide cases in Brazil have been increasing every year, in 2010 the rate of deaths per 100,000 inhabitants was 5.0 in 2019 it jumped to 6.4. The South region has the highest rate of deaths by suicide, in 2019 it was 10.6 per 100,000 inhabitants. The analysis of the results also shows a strong relationship between the underlying cause of suicide with the variables age and state.

**Keywords:** Data Analysis, Suicide, KDD, Public Policy.



## LISTA DE FIGURAS

Figura 1 – Etapas da descoberta de conhecimento em banco de dados – KDD.....	4
Figura 2 – Número de morte por suicídio no Brasil de acordo com o sexo.....	14
Figura 3 – Incidência de suicídio no Brasil.....	16
Figura 4 – Incidência de suicídio por região.....	17
Figura 5 – Número de mortes no Brasil conforme a Raça e/ou Cor.....	18
Figura 6 – Número de caso de suicídio no Brasil conforme o Estado Civil.....	19
Figura 7 – Casos de mortes por suicídio no Brasil conforme a escolaridade.....	20
Figura 8 – Média de idade das pessoas que cometeram suicídio no Brasil.....	21
Figura 9 – Correlação entre as variáveis estudadas.....	22

## LISTA DE QUADROS

Quadro 1 – Dicionário de dados.....	8
Quadro 2 – Atributos selecionados para a análise da base de dados.....	10
Quadro 3 – Atributos sobre a estimativa da população brasileira.....	10
Quadro 4 – Descrição das categorias sobre suicídio - CID 10.....	12
Quadro 5 – Categoria classificatória para o suicídio - CID 10.....	13

## SUMÁRIO

<b>INTRODUÇÃO</b> .....	1
<b>REVISÃO DE LITERATURA</b> .....	3
<b>METODOLOGIA</b> .....	7
Obtenção de Dados.....	7
Pré-processamento dos dados.....	8
Análise dos dados.....	10
<b>RESULTADOS E DISCUSSÕES</b> .....	12
<b>CONSIDERAÇÕES FINAIS</b> .....	24
<b>REFERÊNCIAS</b> .....	26

## INTRODUÇÃO

A tecnologia computacional traz várias possibilidades de aplicação em diversas áreas do conhecimento. É uma ferramenta que tem proporcionado evolução nas distintas áreas científicas, aliando conhecimento especialista com os vários produtos da computação.

A possibilidade de armazenar grande quantidade de dados e o aumento na capacidade de processamento são alguns dos avanços da computação moderna. Fato previsto por Gordon Moore na conhecida *Lei de Moore*, que previa que ao longo do tempo o aumento do desempenho dos computadores dobraria em um período aproximado de dois anos (SILVA *et al*, 2016).

A facilidade para armazenar um volume maior de informação nos computadores mudou a maneira das pessoas guardarem dados, sejam eles pessoais, de grandes empresas privadas ou órgãos públicos.

A adesão e uso de computadores, cada vez mais comum em escolas, universidades, empresas, hospitais, etc., proporcionou um salto de qualidade no registro, guarda e recuperação de dados. O que antes eram documentos escritos e/ou datilografados em papel como fichas, formulários, boletins, relatórios, etc., atualmente são armazenados em pequenas unidades de armazenamento, passível de ser reconhecida e lida por um computador.

Os dados armazenados contêm informações ocultas que, quando reveladas, são de grande importância para a tomada de decisão (CORCOVIA, 2019).

Conforme foi mencionado no parágrafo anterior, o armazenamento de dados em computador traz um potencial enorme de olhar para o passado e compreender, com base nas informações, características até então não reveladas dos dados brutos armazenados, pois facilita a recuperação e análise desses dados.

Uma das principais etapas da descoberta de conhecimento em banco de dados é a mineração de dados, que faz uso de técnicas, algoritmos e ferramentas para a obtenção do conhecimento. A mineração de dados é o processo de encontrar padrões e correlações em grandes conjuntos de dados para prever resultados (MARQUES *et al*, 2020).

Este estudo visa obter conhecimento sobre o suicídio no Brasil. Para isto fará uso de uma base de dados obtida na plataforma kaggle, a base de dados foi

extraída do DATASUS, sistema de dados do Ministério da Saúde que está vinculado à Secretaria de Gestão Estratégica e Participativa, que tem a responsabilidade de coletar, processar e disseminar informações sobre a saúde no Brasil.

Compreender casos de suicídio por meio da obtenção de conhecimento em base de dados pode ajudar o poder público na construção de políticas públicas cada vez mais eficazes, podendo contribuir para a prevenção de tentativas de suicídio.

A mineração de dados é uma técnica que possibilita, através da análise dos dados, a racionalização destes para obter informação que posteriormente será transformada em conhecimento (MARQUES *et al*, 2020).

Diante do exposto e por meio dos dados sobre suicídio, pretende-se extrair conhecimento que possa ajudar os gestores da área da saúde na melhoria das políticas que atuem diretamente na prevenção do suicídio. Tem como objetivo realizar levantamento de dados sobre o suicídio; verificar, através da técnica de mineração e análise de dados, possíveis relações existentes entre os dados; identificar padrões que possam ser úteis para entendimento dos fatores que levam ao suicídio e obter conhecimento útil para compreender a problemática estudada.

## REVISÃO DE LITERATURA

O desenvolvimento científico propicia as inovações tecnológicas, que por sua vez podem contribuir para novas descobertas da ciência. É comum, entretanto, o uso de artefatos tecnológicos em diversas áreas do conhecimento, inclusive na área denominada ciências da saúde.

O uso de tecnologia na área da saúde tem sido fomentado pelo estado brasileiro, que em 2004 aprovou a Política Nacional de Ciência Tecnologia e Inovação em Saúde (PNCTS), tal iniciativa tinha como objetivo incorporar tecnologias nos sistemas de saúde (NOVAES; SOÁREZ, 2020).

Para Barra (2006) a tecnologia atual que foi criada a serviço do homem, vem ajudando na resolução de problemas, na área da saúde, que antes eram insolúveis e, segundo o autor, essa aplicação da tecnologia tem contribuído para melhorar a saúde e a vida das pessoas.

O uso de tecnologia na área da saúde pode ser importante para adoção de medidas de prevenção. Por meio do uso de técnicas específicas é possível descobrir conhecimentos em base de dados e, a partir dos conhecimentos obtidos, racionalizar o planejamento de ações de profilaxia nas diversas áreas da saúde.

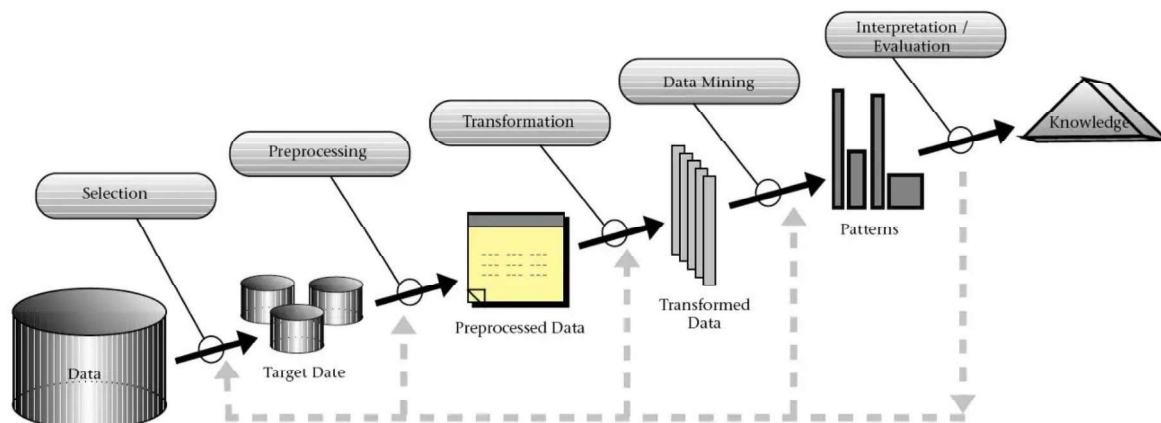
A descoberta de conhecimento em banco de dados – *KDD*, de acordo com Escobar *et al* (2019) é “um conjunto de atividades organizadas em três grandes etapas: pré-processamento, mineração e pós-processamento”, conforme mostrado na Figura 1.

Segundo Macris (2021) o pré-processamento é a retirada de dados que pode interferir na análise (ruídos), a definição de estratégias para o tratamento dos dados faltantes (NaN) e a checagem das informações dos campos. É também uma etapa para o conhecimento dos dados, a forma em que estão distribuídos para facilitar a obtenção das informações pretendidas.

A mineração busca “por padrões de interesse em uma forma representacional específica ou um conjunto de tais representações, incluindo regras de classificação, regressão e agrupamento” (MACRIS, 2021 p. 29).

O pós-processamento visa verificar a importância e qualidade dos dados minerados para posterior apresentação, caso tenha obtido o conhecimento desejado dos dados. Caso os dados obtidos não sejam interessantes é possível realizar novo

processo de mineração com alteração nas bases de preparação, a fim de obter resultados mais interessantes (DA SILVA, 2020).



**Figura 1 - Etapas da descoberta de conhecimento em banco de dados – KDD.**  
Fonte: Han (2012).

Entre as etapas de *KDD* a mais conhecida é a mineração de dados, pois é nessa etapa que os dados são transformados em informação. Porém, todas as etapas são importantes para a qualidade do conhecimento que será extraído ao fim do processo. A falha em alguma das etapas pode induzir a erros e comprometer a interpretação dos resultados.

De acordo com Han *et al.* (2012), a mineração de dados consiste em uma evolução natural da tecnologia e permite transformar um conjunto de dados em conhecimento valioso.

Sendo uma metodologia usada na extração de conhecimento em base de dados, o KDD torna-se um instrumento importante para analisar dados na área da saúde, podendo trazer benefícios ao ser humano ao realizar estudos com a junção de áreas distintas (KACUTA, 2021).

Para a descoberta de conhecimento em base de dados é imprescindível ter dados validados e estruturados e que possibilite o uso da mineração desses dados. Os registros sobre dados de saúde do Brasil podem ser obtidos por meio de plataformas como o kaggle ou no DATASUS.

No Brasil, de acordo com Saldanha *et al* (2019), “após a *Constituição Federal* de 1988 ter estabelecido o Sistema Único de Saúde (SUS), foi criado o Departamento de Informática do SUS (DATASUS) em 1991 visando à coleta e organização de dados referentes ao SUS”.

Os sistemas de informação em saúde mantidos pelo DATASUS, ou em colaboração com ele, cobrem diversos aspectos da saúde populacional. Alguns são de natureza epidemiológica, como o Sistema de Informações sobre Mortalidade (SIM) e o Sistema de Informações sobre Nascidos Vivos (SINASC), que utilizam dados dos cartórios. Já outros sistemas têm objetivos administrativos, como o Sistema de Internações Hospitalares (SIH) e o Sistema de Informações Ambulatoriais (SIA), usando dados provenientes diretamente da assistência à saúde. Mesmo os sistemas de informação de origem administrativa e financeira contêm dados relevantes acerca da situação de saúde brasileira<sup>3</sup> (SALDANHA *et al*, 2019, p. 2).

No DATASUS é possível obter uma grande quantidade de dados, registrados ao longo do tempo, sobre temas diversos que podem ser usados para obtenção de conhecimento.

Segundo Santos (2021), “o DATASUS possui características de um “*Big Data*” e disponibiliza dados sobre a população, internações hospitalares, mortalidade, nascidos vivos e de produção ambulatorial”.

Entre os dados disponibilizados pelo DATASUS é de interesse deste estudo a base de dados sobre mortalidade, especificamente aquela relativa às mortes por suicídio. Compreender melhor este tipo de mortalidade pode ajudar a evidenciar ações para melhoria de políticas públicas em saúde mental.

De acordo com Kacuta (2021) em relatório de estimativas globais de saúde da Organização Mundial de Saúde de 2019, aproximadamente 800 mil pessoas morrem anualmente em decorrência de suicídio no mundo. Por mais que a quantidade de mortes seja alarmante, é baixo o número de países que adotam estratégias para enfrentamento do problema.

O suicídio é um problema de saúde pública multicausal, isto é, não há uma única causa. Quase sempre há algum dos sintomas de transtornos mentais comuns (TMC). Os TMCs, estão relacionados ao sofrimento psíquico e inclui estado de ansiedade, depressão e sintomas de somatização, podem ser manifestados em conjunto ou separadamente. Ainda dentro dos TMC estão inclusos os sintomas de insônia, baixa concentração, falha de memória, fadiga, irritação, sensação de inutilidade, entre outras queixas (QUADROS, 2018).

Entre os sintomas que podem aumentar a chance de suicídio inclui depressão, dependência química, ansiedade grave, crises de pânico, agitação e insônia (Chachamovich *et al*, 2009 *apud* KACUTA, 2021).

Calixto Filho e Zerbini (2016), afirmam que outros fatores também se relacionam com o suicídio, já que é um fenômeno humano multifacetado que depende da interação biológica, genética, psicológica, sociocultural e econômica.



O suicídio não é um problema de saúde que envolve apenas quem tenta tirar a própria vida ou daqueles que têm êxito gerando mais um número para as estatísticas. Afeta todos que fazem parte do ciclo de relações da pessoa envolvida, pois desestabiliza a 'normalidade patológica' vivida pelos contemporâneos que se põe a questionar ou julgar tal atitude, fato totalmente compreensível, dada a pouca ou nenhuma informação que as pessoas detêm sobre sofrimento psíquico, que na cultura popular são classificados como 'frescura', 'coisa de gente que não tem o que fazer', 'falta de trabalho', etc.

A perda humana e o impacto social do suicídio têm consequências materiais e psicológicas para familiares, amigos e demais indivíduos da rede de relações pessoais do morto (MACHADO; SANTOS, 2015).

Culturalmente herdamos o pensamento da filosofia materialista que incorporado ao sistema econômico nos remete à valorização daquilo que é material, concreto. Um exemplo em saúde é a medicina ortopédica (obviamente é muito importante), que na maioria das vezes trata pacientes cuja ferida está ali, exposta, materializada dando aos que convivem com a pessoa a noção de gravidade, mas não é assim quando tratamos de transtornos mentais, seja ele qual for.

A dor da mente não é vista pelos olhos, ela está enclausurada em algum lugar. Lugar este que é sombrio para aquele que está 'preso', necessitando de alguma luz para tentar se safar e salvar a si mesmo.

Quando uma pessoa entende que o caminho para superar algum tipo de sofrimento é tirar a própria vida, é porque as medidas de prevenção não existem ou as que existem falharam e precisam de melhoria. Para que tais melhorias aconteçam é preciso conhecer melhor o fenômeno da morte por suicídio, essa é a proposta da presente pesquisa.

## **METODOLOGIA**

### **Obtenção de Dados**

A base de dados seria obtida diretamente no Sistema de Informação sobre Mortalidade (SIM) do DATASUS, porém em abril e maio do ano de 2022 todas as tentativas para a obtenção da base de dados retornavam um erro na página do Ministério da Saúde do Brasil. Após várias tentativas em diferentes computadores e navegadores, não foi possível obter os dados. Foi enviado um e-mail para a área de suporte informando o problema, porém não houve nenhuma resposta.

Diante da dificuldade com a obtenção dos dados diretamente no DATASUS foi feita uma pesquisa no site kaggle.com, no qual foi possível obter a base de dados do presente estudo<sup>1</sup>. O kaggle é uma plataforma vinculada ao Google que possui uma comunidade online de cientistas de dados.

Outra base de dados utilizada foi a do Instituto Brasileiro de Geografia e Estatística (IBGE), contendo o número da população brasileira entre os períodos de 2010 a 2019. Foi selecionado esse período porque eram escassas, no início do ano de 2022, as informações sobre suicídio no Brasil após o ano de 2019 e o IBGE ainda não havia publicado as estimativas da população brasileira do último ano, devido ao atraso na pesquisa do censo, que é feita a cada dez anos. Outro aspecto importante é que em 2020 teve início a pandemia do coronavírus, fato que poderia causar ruído, caso fossem incluídos os dados desse período no presente estudo.

Foi realizado o download do conjunto de dados no formato CSV (Comma separated values) e posteriormente incluído no Google Drive para posterior análise no recurso disponível da Google, denominado Colaboratory ou Colab<sup>2</sup>, uma ferramenta que permite o uso da plataforma para escrever e executar códigos em Python. Além da facilidade na manipulação de código e uso de bibliotecas sem a necessidade de instalação local, o Colab pode ser utilizado em qualquer computador com acesso à internet, permitindo maior produtividade no uso da ferramenta, já que o arquivo é salvo no Google Drive e pode continuar a ser editado em outra máquina (GOOGLE, 2022).

---

<sup>1</sup><https://www.kaggle.com/datasets/psicodata/dados-de-suicidios-entre-2010-e-2019?resource=download>

<sup>2</sup> <https://colab.research.google.com/drive/1mH14NVuWNt4nn6HwxBp1kVHIZwRsRqN?usp=sharing>

## Pré-processamento dos dados

A base de dados obtida possui um total de 112.491 registros e 18 atributos dos quais, após o pré-processamento, restaram oito. Adicionalmente, foi criado o atributo idade, totalizando nove atributos. Foi necessária a criação do atributo idade para caracterizar melhor os indivíduos que cometeram suicídio.

No quadro 1 é apresentado o dicionário dos dados da base de dados original sobre o suicídio.

### Quadro 1 - Dicionário de dados.

Atributo	Descrição
Seq	Índice
Estado	Estado onde o óbito ocorreu.
Ano	Ano do óbito.
mês	Mês do óbito.
DTOBITO	Data do óbito.
DTNASC	Data do nascimento do falecido.
SEXO	Sexo do falecido.
RACACOR	Cor informada pelo responsável pelas informações do falecido.
ASSITMED	Se refere ao atendimento médico continuado que o paciente recebeu, ou não, durante a enfermidade que ocasionou o óbito.
ESMAE	Escolaridade da mãe em anos.
ESTCIV	Situação conjugal do falecido informada pelos familiares.
ESC	Escolaridade em anos.
OCUP	Tipo de trabalho que o falecido desenvolveu na maior parte de sua vida produtiva.
CODMUNRES	Código do município de residência.
CAUSABAS	Causa básica da DO (definição do óbito). (Códigos CID 10).
CAUSABAS_O	Causa básica informada antes da resseleção. (Códigos CID 10).
LOCOCOR	Local de ocorrência do óbito.
CIRURGIA	Realização de cirurgia.

Fonte: BRASIL (2022).

O critério utilizado para a exclusão dos atributos foi a quantidade de registro vazio (NaN) da base de dados. Assim foi estabelecido o percentual acima de 25% de registros vazios para desconsiderar um atributo. As colunas 'ASSITMED', 'ESMAE',

‘OCUP’ e ‘CIRURGIA’ apresentaram percentual acima de 25% de linhas sem registros (NaN) e foram excluídas do *dataframe*.

As colunas ‘DFOBITO’ e ‘DTNASC’ foram convertidas do formato *object*, formato original na base de dados, para o formato *datetime64* com o objetivo de possibilitar o cálculo para a obtenção da idade da pessoa quando ocorreu a morte por suicídio. Após a conversão de tipos, citada acima, foi criada uma nova coluna no *dataframe* denominada ‘idade’.

Para a obtenção da idade foi feita uma operação matemática básica subtraindo a data de óbito ‘DFOBITO’ pela data de nascimento ‘DTNASC’, o resultado mostra a quantidade de dias vividos. Posteriormente a quantidade de dias foi dividida pelo valor 365 (correspondente a quantidade de dias em um ano), obtendo assim a idade.

Na análise da correlação entre as variáveis foi utilizado o teste de associação qui-quadrado  $\chi^2$ , pois através deste teste é possível realizar análise entre variáveis nominais/categóricas.

O teste de qui-quadrado verifica se as variáveis categóricas possuem independência entre si. De acordo com Anunciação (2021), o valor do  $\chi^2$  é representado pela seguinte equação:

#### **Equação 1 – Qui-quadrado**

$$\chi^2 = \sum_{k=1}^n \frac{(O_k - E_k)^2}{E_k}$$

Segundo Anunciação (2021) as incógnitas possuem as seguintes descrições:

K: é a quantidade de classes.

O: valor observado de uma classe.

E: valor esperado da classe.

Após o pré-processamento, restaram algumas variáveis com pouca significância para a análise dos dados. As variáveis com pouca significância são aquelas que tem pouca interferência na compreensão dos casos sobre o suicídio. Assim foram excluídas as colunas do *dataframe* denominadas ‘seq’, ‘DFOBITO’, ‘DTNASC’, ‘CODMUNRES’, ‘CAUSABAS\_O’ e ‘LOCOCOR’, ao final foi feita a análise dos dados com as informações contidas no Quadro 2.

## Quadro 2 - Atributos selecionados para a análise da base de dados.

Atributo	Tipo de dado	Valores
Estado	Catégorico	27 valores: GO, SP, BA, entre outros.
Ano	Numérico	Valores de 2010 a 2019.
Mês	Numérico	Valores de 1 a 12.
SEXO	Catégorico	Valores: Masculino, feminino.
RACACOR	Catégorico	5 valores: Parda, indígena, entre outros.
ESTCIV	Catégorico	5 valores: Casado/a, Solteiro/a, entre outros.
ESC	Catégorico	5 valores: 1 a 3 anos, 4 a 7 anos, entre outros.
CAUSABAS	Catégorico	25 valores: X70, X74, entre outros.
Idade	Numérico	Valores entre 5 e 112.

A base de dados do IBGE<sup>3</sup> (IBGE, 2022), sobre o número da população brasileira, foi utilizada para verificar a incidência de casos de suicídio ao longo do período temporal analisado, ou seja, entre os anos de 2010 a 2019. Para cada ano havia uma base de dados distinta, portanto após a obtenção de todos os dados, foi feita a junção em uma única base de dados para facilitar a análise. O quadro 3 mostra os atributos contidos no *dataset* do IBGE.

## Quadro 3 - Atributos sobre a estimativa da população brasileira.

Atributo	Tipo de dado	Valores
Brasil e unidades da federação	Catégorico	33 valores: Brasil, Sul, GO, SP, BA, entre outros.
População estimada	Numérico	Valores correspondente ao número da população.

## Análise dos dados

Para trabalhar com os dados, os recursos de software utilizados foram o navegador Google Chrome, Colab e Google Drive.

Para análise dos dados e plotagem dos gráficos foram utilizadas as bibliotecas Pandas, Datetime, Numpy, Math, Scipy, Matplotlib, Seaborn e Scikit-learn da linguagem de programação Python.

---

<sup>3</sup> <https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html?edicao=17283&t=downloads>

Pandas é uma biblioteca utilizada para a manipulação e análise de dados contidos em tabelas ou, conforme a própria biblioteca, *dataframe*; Datetime é uma biblioteca usada para trabalhar com a manipulação de datas e horas; Numpy permite trabalhar com matrizes; Math é uma biblioteca que possibilita trabalhar com funções matemáticas; Scipy é usada na computação científica e facilita na análise estatística e na manipulação de dados; Matplotlib é talvez a mais famosa biblioteca para a elaboração e visualização de gráficos; Seaborn é uma biblioteca usada para plotagem de gráficos estatísticos baseados em um *dataframe*, esta biblioteca possibilita ao usuário criar gráficos sem a necessidade de informar os dados estatísticos prontos, ela possui ferramentas que faz isso para o usuário; a biblioteca Scikit-learn é utilizada para o aprendizado de máquina, com base nos dados é possível fazer uso de regras de associação, regressão, clusterização, etc.

## RESULTADOS E DISCUSSÕES

As causas básicas sobre o suicídio na base de dados original são identificadas pelos códigos do Manual de Classificação Nosológica CID-10. O manual de Classificação Internacional de Doenças (CID-10) é utilizado na área médica para descrever doenças e problemas relacionados à saúde humana.

A classificação para os casos de suicídio no CID-10 é dada pela letra X, seguida de três números. Os dois primeiros números são referentes à categoria e o terceiro número corresponde ao local de ocorrência do suicídio (subcategoria).

Cada categoria se refere a uma modalidade de classificação que identifica o meio empregado para o ato suicida. A descrição de cada categoria é apresentada no Quadro 4.

**Quadro 4 - Descrição das categorias sobre suicídio - CID 10.**

<b>Categoria CID 10</b>	<b>Descrição</b>
<b>X60</b>	Auto-intoxicação por exposição, intencional, a analgésicos, antipiréticos e anti-reumáticos, não-opiáceos.
<b>X61</b>	Auto-intoxicação por exposição, intencional, a drogas anticonvulsivantes [antiepilépticos] sedativos, hipnóticos, antiparkinsonianos e psicotrópicos não classificados em outra parte.
<b>X62</b>	Auto-intoxicação por exposição, intencional, a narcóticos e psicodislépticos [alucinógenos] não classificados em outra parte.
<b>X63</b>	Auto-intoxicação por exposição, intencional, a outras substâncias farmacológicas de ação sobre o sistema nervoso autônomo.
<b>X64</b>	Auto-intoxicação por exposição intencional a outras drogas, medicamentos e substâncias biológicas e às não especificadas.
<b>X65</b>	Auto-intoxicação voluntária por álcool.
<b>X66</b>	Auto-intoxicação intencional por solventes orgânicos, hidrocarbonetos halogenados e seus vapores.
<b>X67</b>	Auto-intoxicação intencional por outros gases e vapores.
<b>X68</b>	Auto-intoxicação por exposição, intencional, a pesticidas.
<b>X69</b>	Auto-intoxicação por exposição, intencional, a outros produtos químicos e substâncias nocivas não especificadas.
<b>X70</b>	Lesão autoprovocada intencionalmente por enforcamento, estrangulamento e sufocação.
<b>X71</b>	Lesão autoprovocada intencionalmente por afogamento e submersão.
<b>X72</b>	Lesão autoprovocada intencionalmente por disparo de arma de fogo de mão.

<b>X73</b>	Lesão autoprovocada intencionalmente por disparo de espingarda, carabina, ou arma de fogo de maior calibre.
<b>X74</b>	Lesão autoprovocada intencionalmente por disparo de outra arma de fogo e de arma de fogo não especificada.
<b>X75</b>	Lesão autoprovocada intencionalmente por dispositivos explosivos.
<b>X76</b>	Lesão autoprovocada intencionalmente pela fumaça, pelo fogo e por chamas.
<b>X77</b>	Lesão autoprovocada intencionalmente por vapor de água, gases ou objetos quentes.
<b>X78</b>	Lesão autoprovocada intencionalmente por objeto cortante ou penetrante.
<b>X79</b>	Lesão autoprovocada intencionalmente por objeto contundente.
<b>X80</b>	Lesão autoprovocada intencionalmente por precipitação de um lugar elevado.
<b>X81</b>	Lesão autoprovocada intencionalmente por precipitação ou permanência diante de um objeto em movimento.
<b>X82</b>	Lesão autoprovocada intencionalmente por impacto de um veículo a motor.
<b>X83</b>	Lesão autoprovocada intencionalmente por outros meios especificados.
<b>X84</b>	Lesão autoprovocada intencionalmente por meios não especificados.

Fonte: CID-10 (2007).

No quadro 5 são mostrados os resultados obtidos na base de dados conforme a categoria de suicídio.

#### Quadro 5 - Categoria classificatória para o suicídio - CID 10.

<b>Categoria CID10</b>	<b>X60</b>	<b>X61</b>	<b>X62</b>	<b>X63</b>	<b>X64</b>	<b>X65</b>	<b>X66</b>	<b>X67</b>	<b>X68</b>	<b>X69</b>
<b>Percentual</b>	0,11%	1,85%	0,57%	0,12%	2,04%	0,42%	0,11%	0,26%	4,50%	2,21%
<b>Categoria CID10</b>	<b>X70</b>	<b>X71</b>	<b>X72</b>	<b>X73</b>	<b>X74</b>	<b>X75</b>	<b>X76</b>	<b>X77</b>	<b>X78</b>	<b>X79</b>
<b>Percentual</b>	67,14%	1,13%	2,47%	0,41%	5,73%	0,05%	1,57%	0,04%	1,61%	0,69%
<b>Categoria CID10</b>	<b>X80</b>	<b>X81</b>	<b>X82</b>	<b>X83</b>	<b>X84</b>					
<b>Percentual</b>	3,79%	0,19%	0,44%	0,23%	2,30%					

A lesão autoprovocada intencionalmente por enforcamento, estrangulamento e sufocação é majoritariamente a principal causa empregada para os atos de suicídio. Disparo de arma de fogo é a segunda causa de morte por suicídio; intoxicação por pesticidas é a terceira causa de morte e a precipitação de um lugar elevado é o quarto meio utilizado para cometimento de suicídio. Com exceção da categoria X70, que é a de maior prevalência, os resultados mostram que os outros três meios empregados para cometimento de suicídio apresentam percentual próximo,



sendo quase sempre um dos três meios o segundo mais usado por pessoas que tiram a própria vida.

As informações obtidas na análise dos dados do *dataframe* mostra que a grande maioria das mortes por suicídio no Brasil são de pessoas do sexo masculino. Em 2010, início da série do presente estudo a relação percentual de mortes por suicídio no Brasil entre homens era de (78%) e (22%) de mulheres, no final da série em 2019 os valores permaneceram inalterados. Na Figura 2 ainda é possível perceber que há uma tendência de crescimento no número de casos de suicídio de homens maior que das mulheres. Tais achados corroboram com os estudos de Machado e Santos (2015) que afirma que “o suicídio é predominante no sexo masculino”, a autora ainda acrescenta, no Brasil, a ocorrência também é expressivamente maior entre homens, corroborando a tendência mundial de que os homens são três vezes mais propensos do que as mulheres a cometer suicídio (MACHADO; SANTOS, 2015).

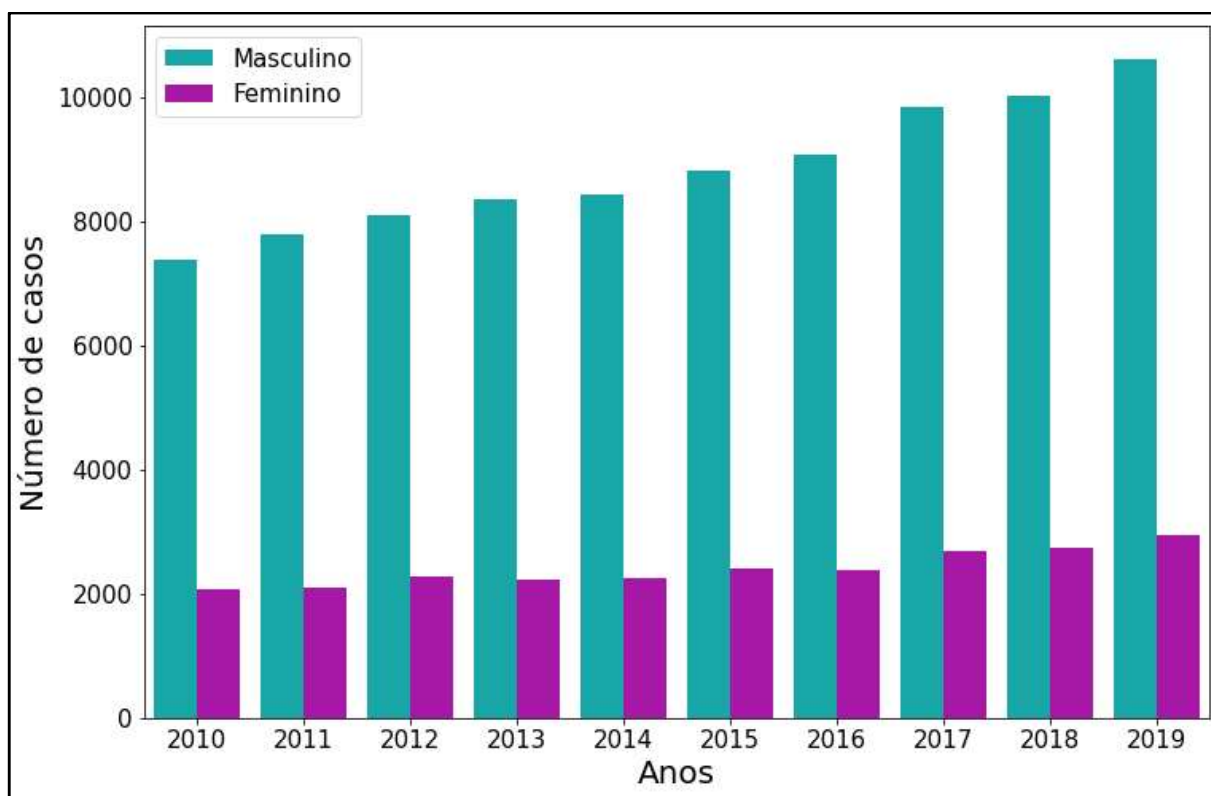


Figura 2 - Número de morte por suicídio no Brasil de acordo com o sexo.

Porém Calixto Filho e Zerbini (2016) salienta que a maior prevalência de tentativas de suicídio foi verificada entre as mulheres, que utilizam métodos de menor letalidade como a intoxicação por medicamentos. Segundo a autora os homens que tentam suicídio utilizam-se de técnicas mais letais, como enforcamento e uso de armas de fogo. Essa tese corrobora com os achados do presente estudo, no qual 83%

dos casos de enforcamento são de pessoas do sexo masculino e 17% do sexo feminino. Em relação ao uso de arma de fogo para cometer suicídio, 90% dos casos são do sexo masculino e 10% feminino.

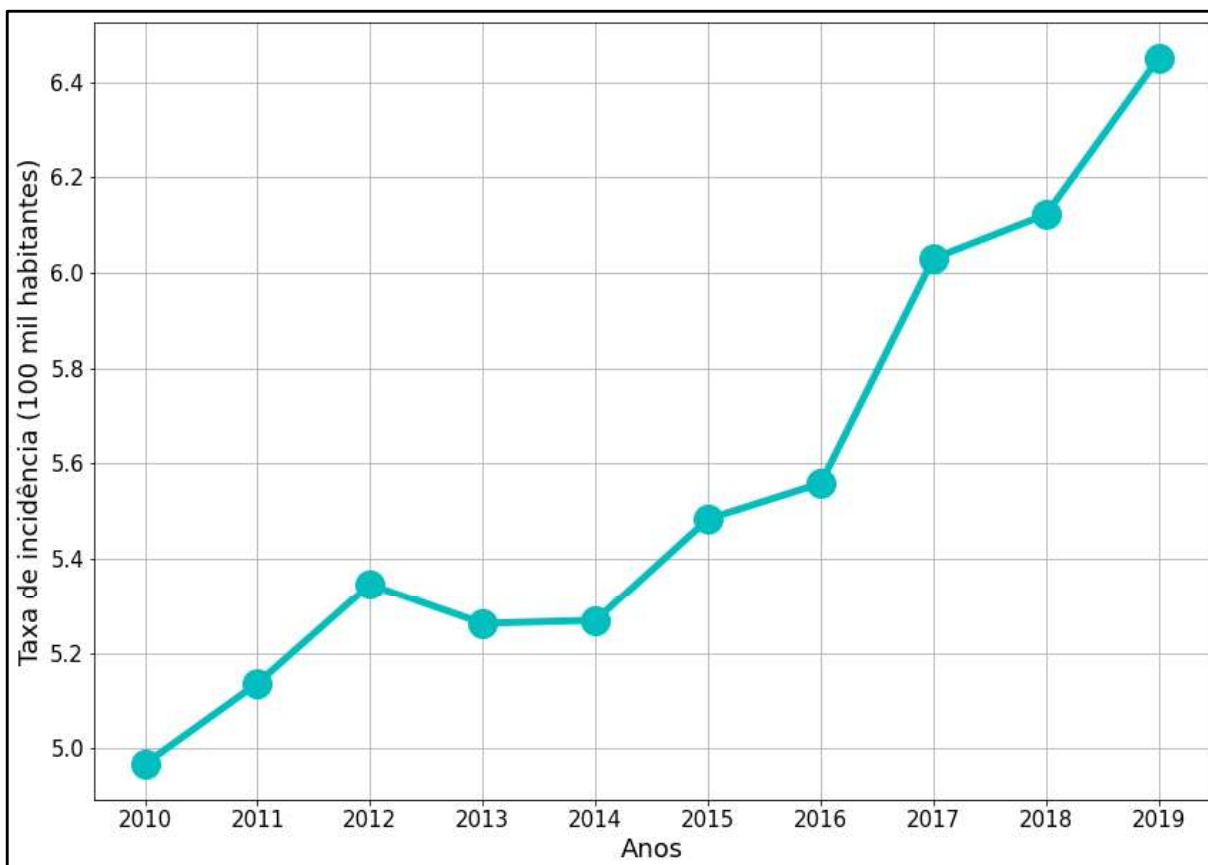
Em uma sociedade machista em que a figura masculina é vista simbolicamente como detentora da força e da coragem, pode contribuir para que as questões relativas à saúde mental sejam negligenciadas, afinal desde pequeno os meninos são encorajados e desafiados a reprimir suas emoções, uma atitude que representa bem essa situação é a célebre frase “homem não chora”, ensinada aos garotos desde pequeno. Falar de sentimento, das emoções é coisa de mulher ou “frescura”, essa é a percepção generalista no mundo masculino, ou seja, todas as portas se fecham quando há algum problema de ordem emocional.

Ademais, as mulheres se preocupam mais com a própria saúde, principalmente mental e emocional, e buscam ajuda profissional com menor resistência que os homens (CALIXTO FILHO; ZERBINI, 2016).

A taxa total de suicídio no Brasil cresceu de 5,0 para 6,4 por 100.000 habitantes, o que representa um crescimento de (28%) no período. No triênio 2012-2014 a taxa permaneceu estável com 5,3 mortes por 100.000 habitantes, porém após este período a taxa vem aumentando ano após ano. Na Figura 3 é mostrada a incidência de casos de suicídio no Brasil.

O número de mortes saltou de 9.476 casos no ano de 2010 para 13.552 no ano de 2019, um aumento bruto de 4.076 novos casos de suicídio. Fato que mostra a necessidade de compreensão da problemática estudada em nível nacional para implementação de uma intervenção mais assertiva e que possa tentar mitigar o comportamento suicida na sociedade brasileira.

Atualmente existem algumas instituições em nível nacional, como os Centros de Atenção Psicossocial (CAPs), o Centro de Referência de Assistência Social (CRAS) e o Centro de Referência Especializado de Assistência Social (CREAS), que podem ajudar a compreender e tratar do tema, porém carecem de recursos e empenho do poder público na promoção de tais instituições.



**Figura 3 - Incidência de suicídio no Brasil.**

Já a análise de casos de suicídio no Brasil por região geográfica, mostra que a região sul é a que possui a maior taxa de mortes, no período estudado a taxa variou entre 7,9 e 10,6 uma variação de (34,17%). É a região que teve a maior incidência de mortes por suicídio no período estudado. Fato que corrobora com os estudos de LOVISI (2009), que estudou o suicídio no Brasil entre 1980 e 2006. A contribuição do autor, demonstra que a região sul se destaca como sendo a região com maior incidência de morte por suicídio no país e a região centro-oeste sendo a segunda maior. Tais resultados são idênticos ao que foi encontrado no presente estudo, conforme apresentado na Figura 4.

A região centro-oeste é a segunda em taxa de incidência com variação entre 5,8 e 7,9 com crescimento percentual de (36,20%), portanto percentualmente superior à variação da região sul no período.

A região sudeste esteve um pouco acima das regiões norte e nordeste, mas em 2018 e em 2019 foi ultrapassada pela região norte. Já a região nordeste esteve, na maior parte do período, abaixo das demais regiões brasileiras em número de incidência de casos de suicídio. A análise de estudos anteriores como os de Lovisi

(2009) e de Calixto Filho e Zerbini (2016) confirma os achados do presente estudo de que as regiões sul e centro-oeste se mantiveram com os maiores índices de suicídio.

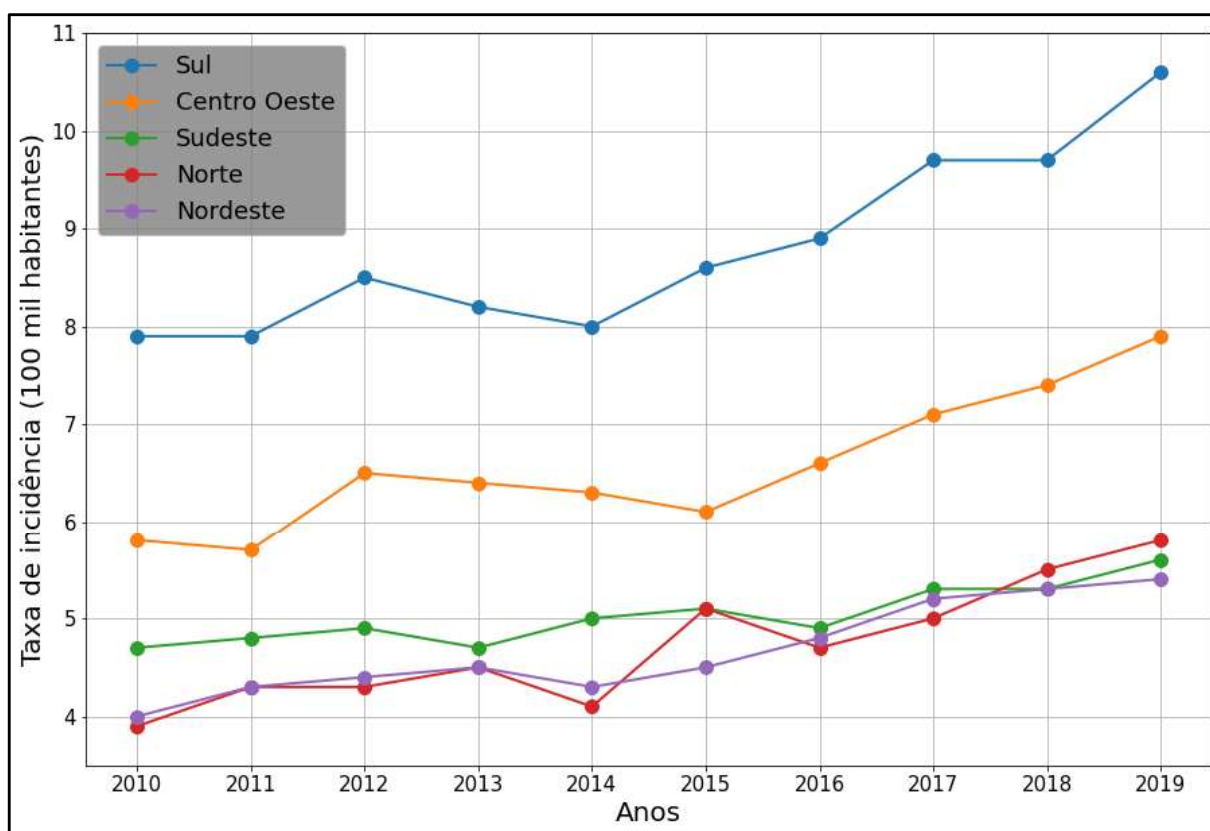
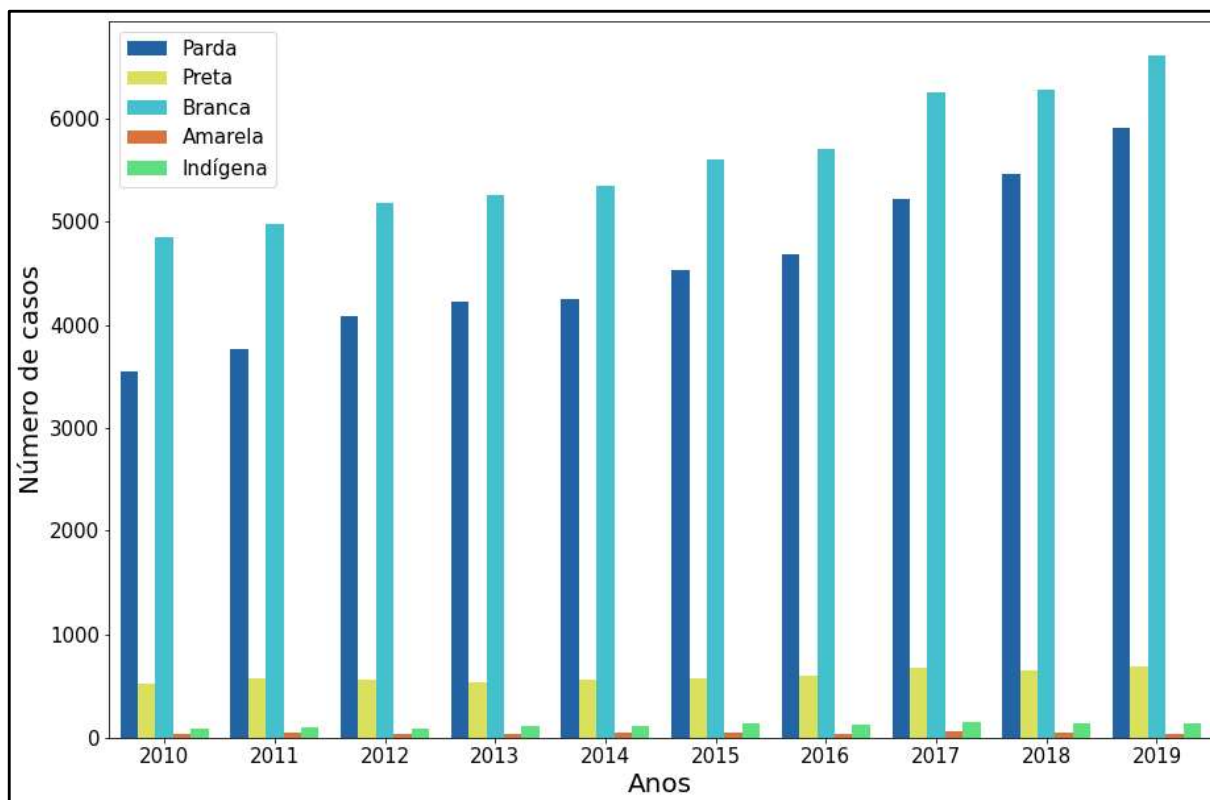


Figura 4 - Incidência de suicídio por região.

A base de dados permitiu também olhar para o estudo do suicídio na ótica da divisão entre raça e cor da pele. A Figura 5 mostra o resultado dos dados brutos relativos às mortes por suicídio de indivíduos, segundo sua Raça/cor.

A maior taxa de mortalidade por suicídio a cada 100 mil habitantes em 2010 foi da população indígena (11,32); seguida da população branca (5,35); parda (4,27), preta (3,61) e por último a amarela com taxa de (1,75) mortes.

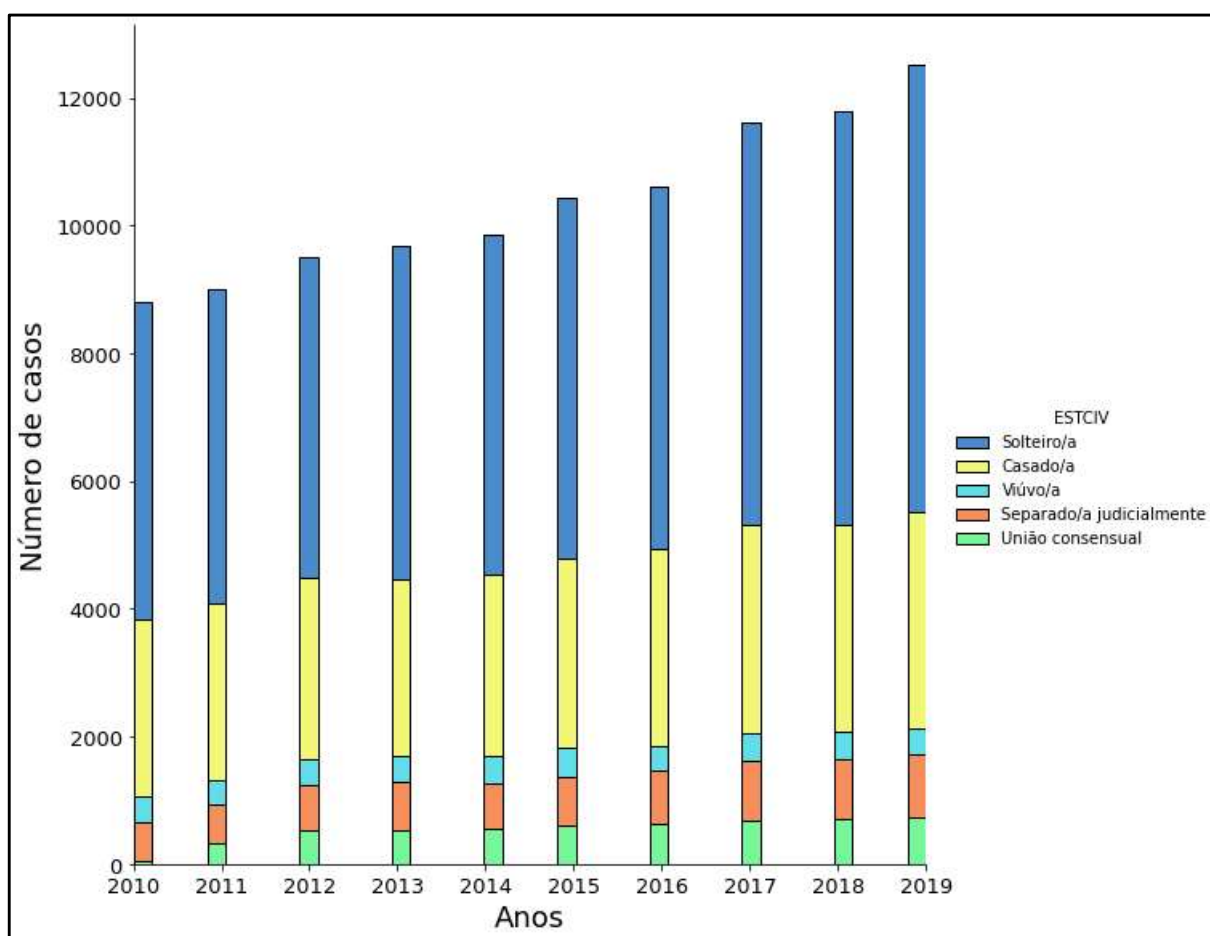
Nos estudos de Machado e Santos (2015) os indígenas também apresentaram a maior taxa de mortalidade, por 100 mil habitantes, no ano 2000 (8,6); seguido pelos amarelos (6,4); brancos (5,4) e pardos (3,3). Já no ano de 2012 as maiores taxas de suicídio, por 100 mil habitantes foram registradas entre os autodeclarados indígenas (14,4); pardos (5,9); os amarelos tiveram redução na taxa de mortalidade por suicídio quando comparada com o ano 2000, apresentando uma taxa de (2,2).



**Figura 5 - Número de mortes no Brasil conforme a Raça e/ou Cor.**

Outro aspecto estudado foi o número de mortes por suicídio relacionado com o estado civil da pessoa na data da morte. A Figura 6 mostra os resultados em valores absolutos para cada estado civil. Considerando apenas o número de casos total, é possível constatar que em 2010 a quantidade de mortes entre os Solteiros foi de 4.400 mortes e no ano de 2019 esse número era de 6.218 mortes. Em 2010 os casados eram 2.491 e em 2019 o número de mortes foi de 3.038. Já os Viúvos somaram 364 mortes em 2010 e 391 no ano de 2019. As pessoas cujo estado civil era Separado Judicialmente, em 2010 totalizaram 554 casos e em 2019 aumentou para 878. Os que tinham União Consensual eram 33 casos de mortes por suicídio no início do período analisado e 661 no final do período.

Nos estudos de Lovisi (2009) o autor afirma que “as principais características sociodemográficas das pessoas que cometeram suicídio durante o período estudado foram baixo nível educacional e estado civil solteiro”.



**Figura 6 - Número de caso de suicídio no Brasil conforme o Estado Civil.**

Em 2000 a Organização Mundial de Saúde (OMS) destacou os seguintes fatores de risco para o suicídio: sexo masculino, faixa etária entre 15 e 44 anos e acima de 75 anos, estratos econômicos extremos, moradores de áreas urbanas, desempregados, aposentados, pessoas em isolamento social, solteiros, viúvos ou separados e migrantes (CALIXTO FILHO; ZERBINI, 2016).

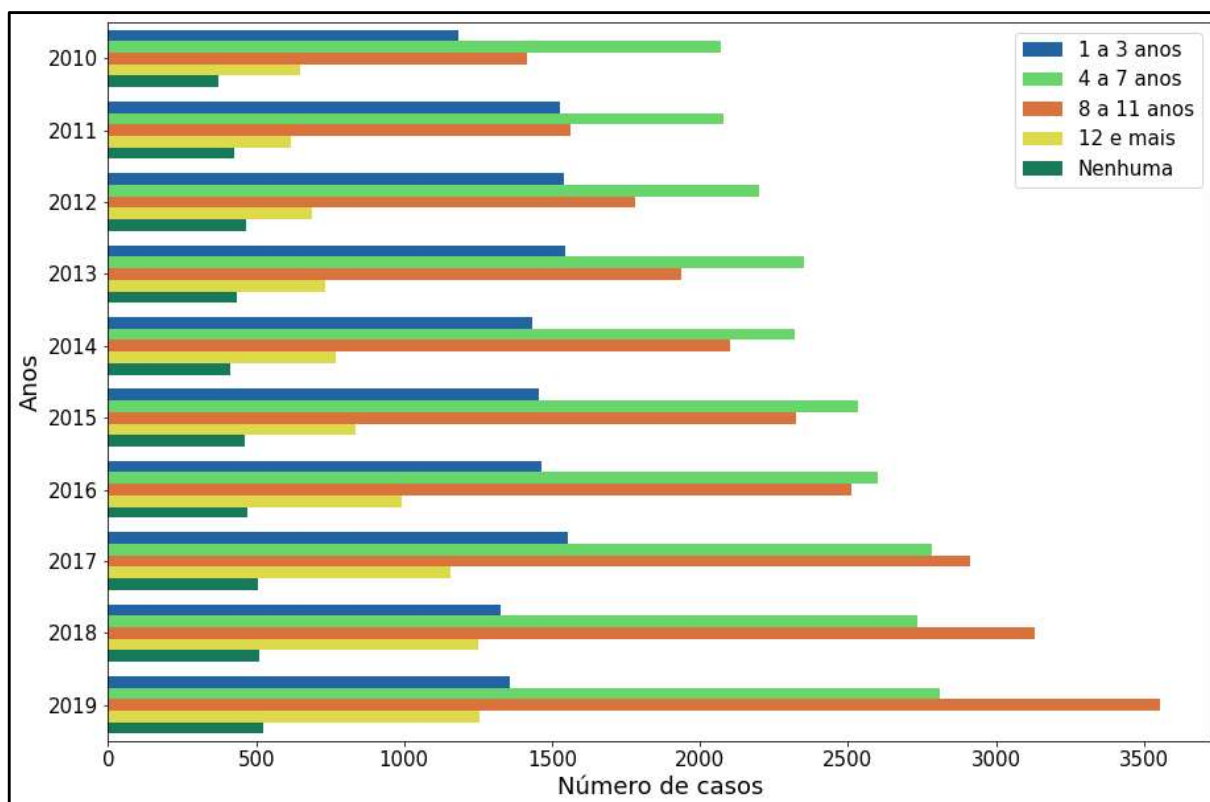
É possível identificar que entre os fatores de risco para o suicídio, elencados pela OMS, alguns estão presentes e são confirmados neste estudo.

A escolaridade também é um dos fatores que devem ser considerados para o risco de suicídio, conforme a OMS. Na Figura 7 são mostrados os resultados obtidos. Os dados estão divididos em períodos trienais, com exceção daqueles que possuíam mais de 12 anos de escolaridade e para os que não tinham nenhuma escolaridade.

Em 2010 o percentual de mortos por suicídio, segundo a escolarização foram os seguintes: (6,53%) para indivíduos sem nenhuma escolaridade; (20,78%) para indivíduos entre 1 a 3 anos de escolaridade; (36,38%) para quem tinha entre 4 e 7 anos de estudo; (24,89%) com escolaridade entre 8 e 11 anos e (11,42%) para os

que tinham mais de 12 anos de formação escolar. No final do período os indivíduos com nenhuma escolaridade tiveram o percentual de (5,52%); escolaridade de 1 a 3 anos (14,29%); com tempo de estudo entre 4 e 7 anos foram (29,57%); de 8 a 11 anos de estudos foi registrado o percentual de (37,40%) e com mais de 12 anos de escolaridade (13,22%).

É possível perceber que nos últimos três anos do período analisado as pessoas com escolaridade de 8 a 11 anos passaram a ser maioria no número de casos de suicídio.



**Figura 7 - Casos de mortes por suicídio no Brasil conforme a escolaridade.**

A média de idade das pessoas que cometeram suicídio é apresentada anualmente na Figura 8. A média de idade global no período de 2010 a 2019 é de 42 anos.

Os resultados dos estudos de Calixto Filho e Zerbini (2016) mostram que os índices de suicídio aumentaram em todas as faixas etárias em todas as pesquisas estudadas. No presente estudo as faixas etárias com maior crescimento se deu entre 20 e 59 anos, com porcentagem que oscilou de 22% a 30% no período analisado.

Portanto, a média de idade das pessoas que cometeram suicídio se encontra dentro da faixa etária em que houve aumento no número de mortes.

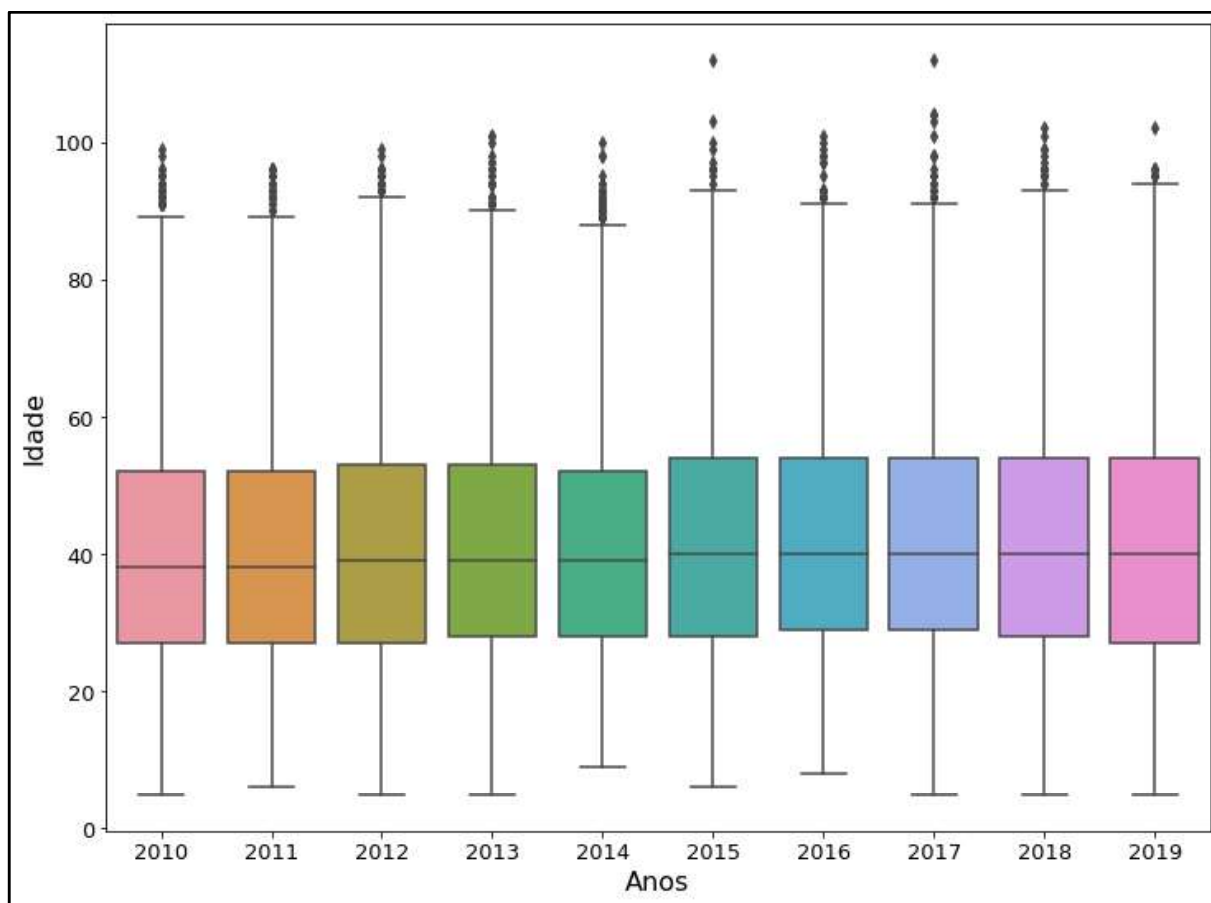


Figura 8 - Média de idade das pessoas que cometeram suicídio no Brasil.

Os dados apresentados mostram que o perfil do suicídio no Brasil tem características que podem ser destacadas em outros estudos com temática idêntica a que foi trabalhada na presente pesquisa. Em quase todos os casos, os dados de um estudo confirmam os achados de outros. Esse é um fato relevante, pois mostra que ao longo do tempo as faces do indivíduo propenso a cometer suicídio está relativamente identificada.

Nos estudos de correlação, entre as variáveis estudadas, a Figura 9 mostra a associação entre as variáveis presentes no *dataframe*.

Os valores de cada correlação variam de -1 (correlação negativa perfeita) a +1 (correlação positiva perfeita), já quando o valor é igual a 0, significa que não há uma correlação aparente entre as variáveis. É possível perceber que há uma relação positiva entre as variáveis **causa básica** (CAUSABAS) com a variável **estado** e **causa básica** com a variável **idade**. Trata-se de uma correlação positiva perfeita cujo valor



é '1', o que significa que há uma forte associação entre essas variáveis e que elas estão diretamente correlacionadas.

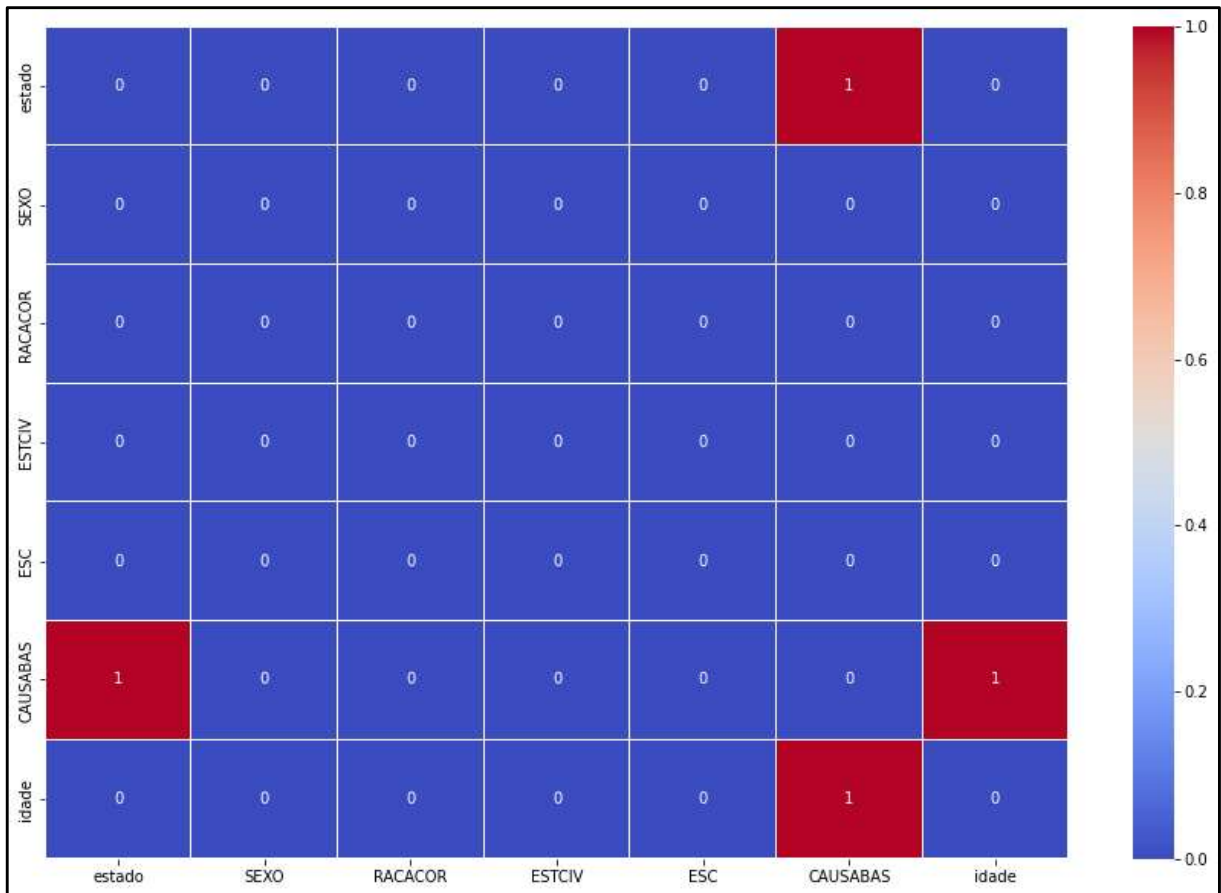


Figura 9 - Correlação entre as variáveis estudadas.

A CAUSABAS é a variável que representa o tipo de lesão empregada no ato suicida, no estudo de correlação é importante compreender melhor a relação entre as variáveis que obtiveram resultados que apontam para a associação. A correlação entre idade e CAUSABAS pode ajudar a compreender se há alguma faixa etária em que o meio empregado para o ato suicida seja mais comumente utilizado. Já em relação a variável estado e CAUSABAS também pode verificar se há algum estado cujo meio empregado no ato suicida seja mais utilizado. Esses são achados importantes, já que a idade da pessoa que comete suicídio tem relação direta com a classificação do tipo de suicídio e o estado também mostra essa relação.

Por fim foi realizado o processo de seleção de variável com o objetivo de identificar as variáveis de maior relevância da base de dados. A seleção de variável ajudou a descobrir os melhores atributos contidos no conjunto de dados usados no presente estudo. Para Silva *et al* (2021) a seleção de variável é um processo que

objetiva a redução do número de variáveis e visa encontrar quais das variáveis são mais significativas.

As variáveis categóricas com maior significância foram respectivamente SEXO, RACACOR, ESTCIV e por último ESC. O que mostra que há uma associação estatisticamente significativa entre essas variáveis e a causa de suicídio.

Destacada como uma variável significativa, SEXO é uma variável fundamental e a análise de dados mostra que há relevância para a compreensão das mortes por suicídio, já que a maior parte das pessoas que morrem desta causa são do sexo masculino; as variáveis RACACOR, ESTCIV e ESC apresentam algumas tendências importantes que foram destacadas na discussão dos dados e nos resultados acima.

## CONSIDERAÇÕES FINAIS

Anualmente milhares de pessoas desistem de continuar a viver, por diversos fatores, a desesperança e a incapacidade de lidar com as contingências do cotidiano as leva para um caminho em que a solução, vista por elas, é tirar a própria vida.

Identificar características do suicídio pode ajudar na compreensão desse problema de saúde, que gera milhares de mortes anualmente no Brasil e no mundo.

As técnicas de análise de dados estão cada vez mais sofisticadas, fato que potencializa seu uso para aplicação e descoberta de conhecimento em diversas áreas do conhecimento, inclusive a área da saúde que possui, no DATASUS, uma variedade de sistemas de registros de dados sobre saúde.

Considerando as premissas descritas acima, é possível obter informações em base de dados que seja útil para a proposta de políticas públicas de combate ao suicídio no Brasil, conforme apresentado na seção de resultados e discussão acima, há uma rica fonte de informação em base de dados que contribui para a compreensão de diversos aspectos relacionados ao suicídio, desde aspectos macrossociais que visa compreender o suicídio no País, nas grandes regiões geográficas, nos estados e municípios, que não foram explorados no presente estudo, a elementos mais subjetivos como a escolaridade, Raça/cor, por exemplo.

Os dados mostraram que há um crescimento no número de casos de suicídio no Brasil. No período estudado a taxa subiu de 5,0 para 6,4 mortes de pessoas por 100 mil habitantes. Quando a análise é feita entre as grandes regiões geográficas, as regiões Sul e Centro-oeste lideram, respectivamente, a taxa de mortes por suicídio. É possível destacar ainda que a região Centro-oeste teve um crescimento percentualmente maior (36,20%), contra (34,17%) da região Sul no período estudado. Quando se trata da escolaridade houve uma alteração entre o período estudado, em 2010 as pessoas com escolaridade de 4 a 7 anos eram as que tinham maior percentual de casos de suicídio (36,38%), em 2019 as pessoas com escolaridade entre 8 a 11 anos eram maioria com (37,40%) das mortes. Em relação a idade, 42 anos é a média de idade das pessoas que morreram em decorrência do suicídio.

Vale salientar que na base de dados utilizada alguns atributos tiveram que ser excluídos devido a falta de dados, atributos importantes para a compreensão do

problema de estudo proposto. Portanto, quanto melhor for a qualidade dos dados coletados, maior serão as possibilidades de compreensão do fenômeno estudado.

Conclui se que há muito o que ser pesquisado sobre a temática desenvolvida no presente estudo, como a relação entre a causa básica do suicídio com o estado onde a pessoa reside e sua idade, uma melhor compreensão sobre a alteração do nível de escolaridade no perfil de pessoas que cometem suicídio também são temas a serem explorados em estudos futuros. Porém, existem estudos consistentes que podem ser utilizados como referência para a pesquisa sobre o suicídio no Brasil. A riqueza existente entre os dados são fontes fundamentais para a compreensão das problemáticas de saúde no país.

Apesar de serem tratadas como números no presente estudo, cada pessoa representada na base de dados tinha seus sonhos, famílias, amigos e por algum motivo decidiu dar fim à própria vida. Com o uso de artefatos tecnológicos, este trabalho é um singelo esforço na tentativa de entender tais motivos e contribuir para a compreensão do problema, podendo assim ajudar a evitar o mesmo fim para outras pessoas.

## REFERÊNCIAS

ANUNCIÇÃO, L. **Conceitos e análises estatísticas com R e JASP**. Ed. Beta 2. PUC-Rio, 2021. Disponível em: <<https://bookdown.org/luisfca/docs/>> Acesso em: 10 ago. 2022.

BARRA, D.C.C. *et al.* **Evolução histórica e impacto da tecnologia na área da saúde e da enfermagem**. Revista Eletrônica de Enfermagem, v. 08, n. 03, p. 422 – 430, 2006. Disponível em: <https://www.revistas.ufg.br/fen/article/view/7081/5012> Acesso em: 08 mai. 2022.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. **Estrutura do SIM**. Brasília, 2022. Disponível em: <[https://opendatasus.saude.gov.br/pt\\_BR/dataset/sim-2020-2021](https://opendatasus.saude.gov.br/pt_BR/dataset/sim-2020-2021)> Acesso em: 27 ago. 2022.

CALIXTO FILHO, M.; ZERBINI, T. **Epidemiologia do suicídio no Brasil entre os anos de 2000 e 2010**. Saúde Ética & Justiça, 21(2), 45-51, 2016. Disponível em <<https://doi.org/10.11606/issn.2317-2770.v21i2p45-51>> Acesso em: 15 mai. 2022.

CID-10. Organização Mundial da Saúde. **Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde**. Tradução: Centro Colaborador da OMS para a Classificação de Doenças em Português. 10. ed. Ver. São Paulo: EDUSP, 2007.

CORCOVIA, L. O.; ALVES, R. S. **Aprendizagem de máquina e mineração de dados: avaliação de métodos de aprendizagem**. Revista Interface Tecnológica, V. 16, n. 1, p. 90-101, 2019. Disponível em: <[https://revista.fatectq.edu.br/interfacetecnologica/article /view/562](https://revista.fatectq.edu.br/interfacetecnologica/article/view/562)> Acesso em: 25 ago. 2022.

DA SILVA, L. R. A. B. **Inteligência artificial em processos de extração de conhecimento KDD e KDT**. Revista de Estudos Universitários - REU, Sorocaba, SP, v. 46, n. 1, p. 161–180, 2020. DOI: 10.22484/2177-5788.2020v46n1p161-180. Disponível em: <https://periodicos.uniso.br/reu/article/view/3694>. Acesso em: 25 set. 2022.

ESCOBAR, L. F. A. *et al.* **Descoberta de padrões para identificação de casos de alto custo em operadoras de planos de saúde**. Revista Stricto Sensu, v. 04, n. 01, jan./jun., p. 01-21, 2019. Disponível em: <<http://revistastrictosensu.com.br/ojs/index.php/rss/arti cle/viewFile /57/59>> Acesso em: 08 mai. 2022.

GOOGLE. 2022. Ceres.: Google Drive. <https://drive.google.com>

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. **Data Mining. Concepts and Techniques**. Morgan Kaufmann Publishers is an imprint of Elsevier. 2012.

IBGE – Instituto Brasileiro de Geografia e Estatística. **Estimativas da População**. Rio de Janeiro: IBGE, 2022. Disponível em: <<https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html?edicao=17283&t=downloads>> Acesso em: 25 set. 2022.

KACUTA, S. L. **Utilização da metodologia KDD para descoberta de conhecimento em dados relacionados a toxicodependência e ideação suicida**. 2021. Disponível em: <<https://riut.utfpr.edu.br/jspui/bitstream/1/27606/1/kddtoxicodependenciaideacaosuicida.pdf>> Acesso em: 07 mai. 2022.

LOVISI, G. M. *et al.* **Análise epidemiológica do suicídio no Brasil entre 1980 e 2006**. Rev. Bras. Psiquiatria, 2009. Disponível em <<https://www.scielo.br/j/rbp/a/x7987JHsK6HpNdZn9qkrVtQ/abstract/?lang=pt>> Acesso em: 23 jul. 2022.

MACHADO, D. B.; SANTOS, D. N dos. **Suicídio no Brasil, de 2000 a 2012**. Jornal Brasileiro de Psiquiatria, 2015. Disponível em: <<https://doi.org/10.1590/0047-2085000000056>> Acesso em: 25 set. 2022.

MACRIS, A. H. **Estudo de modelos para predição de reinternação em saúde mental no Sistema Único de Saúde (SUS)**. 2021. Disponível em: <<https://www.teses.usp.br/teses/disponiveis/82/82131/tde-07072021-153520/en.php>> Acesso em: 07 mai. 2022.

MARQUES, R. D. R. DIAS, J. C. DIAS, J.C. **Relevância dos sistemas de informação e mineração de dados no sistema de saúde: urgência e emergência**. Revista Processando o Saber - ISSN 2179-5150, Ano 12, nº 12, 2020. Disponível em: <<https://www.fatecpg.edu.br/revista/index.php/ps/article/download/209/123>> Acesso em: 15 jul. 2022.

NOVAES, H. M. D.; SOÁREZ, P. C. de. **A Avaliação das Tecnologia em Saúde: origem, desenvolvimento e desafios atuais. Panorama internacional e Brasil**. Faculdade de medicina, Universidade de São Paulo, São Paulo, Brasil. Cadernos de Saúde Pública, 2020. Disponível em: <<https://www.scielo.br/j/csp/a/6p3SzRQKCpcR678Btk5xVvYQ/?lang=pt>> Acesso em: 28 ago. 2022.

QUADROS, L de C. M de. **Transtornos mentais comuns e fatores contemporâneos: coorte de nascimentos de 1982**. Revista Brasileira de Enfermagem - REBEn, 2018. Disponível em: <<https://www.scielo.br/j/reben/a/Z7gbHWDxJTKK8jpJKnZrMQx/?format=pdf&lang=pt>> Acesso em: 08 mai. 2022.

SALDANHA, R. de F. BASTOS, R. R. BARCELLOS, C. **Microdatasus: pacote para download e pré-processamento de microdados do Departamento de Informática do SUS (DATASUS)**. Cadernos de Saúde Pública, 2019. Disponível em:

<<https://www.scielo.br /j/csp/a/gdJXqcrW5PPDHX8rwPDYL7F/?lang=pt>> Acesso em: 06 mai. 2022.

SANTOS, D. J. da S. **Extração de conhecimento sobre internações de atenção básica a partir de dados do DATASUS na região metropolitana de São Paulo.** 2021. Disponível em: <<http://ric.cps.sp.gov.br/handle/123456789/8060>> Acesso em: 06 mai. 2022.

SILVA, F. A. B. da. **Big Data e Nuvens Computacionais: Aplicações em Saúde Pública e Genômica.** 2016. Disponível em: <<https://jhi.sbis.org.br/index.php/jhi-sbis/article/view/336 /263>> Acesso em: 12 jul. 2022.

SILVA, M. E. A.; HOLANDA, V. G. L.; SILVA, R. S. da; SEVERIANO, P. V. L.; SILVA, R. de A. **Seleção de Características Biológicas para Prognóstico de Câncer: Revisão Sistemática da Literatura.** Journal of Health Informatics, Brasil, v. 12, 2021. Disponível em: <https://jhi.sbis.org.br/index.php/jhi-sbis/article/view/817>. Acesso em: 27 set. 2022.