

Alana Baldez de Avila

**Violência de Gênero Durante a Pandemia:
Identificação de Padrões em Notificações do
Rio Grande do Sul Usando Técnicas de
Agrupamento de Dados Mistos**

Rio Grande, Rio Grande do Sul, Brasil

Janeiro, 2025

Alana Baldez de Avila

**Violência de Gênero Durante a Pandemia: Identificação
de Padrões em Notificações do Rio Grande do Sul
Usando Técnicas de Agrupamento de Dados Mistos**

Trabalho de Conclusão de Curso apresentado
ao Instituto de Matemática, Estatística e Física da Universidade Federal do Rio Grande
como requisito parcial para a conclusão do
curso de Matemática Aplicada Bacharelado.

Universidade Federal do Rio Grande - FURG

Instituto de Matemática, Estatística e Física - IMEF

Curso de Matemática Aplicada Bacharelado

Orientador: Professor Doutor Juliano César Marangoni

Coorientador: Professora Doutora Raquel da Fontoura Nicolette

Rio Grande, Rio Grande do Sul, Brasil

Janeiro, 2025

Alana Baldez de Avila

Violência de Gênero Durante a Pandemia: Identificação de Padrões em Notificações do Rio Grande do Sul Usando Técnicas de Agrupamento de Dados Mistos

Trabalho aprovado.

Rio Grande, 03 de Fervereiro de 2025.

Documento assinado digitalmente
 JULIANO CESAR MARANGONI
Data: 07/02/2025 10:34:29-0300
Verifique em <https://validar.iti.gov.br>

**Prof^o Doutor Juliano César
Marangoni**

 Documento assinado digitalmente
RAQUEL DA FONTOURA NICOLETTE
Data: 12/02/2025 13:56:01-0300
Verifique em <https://validar.iti.gov.br>

**Prof^a Doutora Raquel da Fontoura
Nicolette**

(Coorientadora IMEF - FURG)
Documento assinado digitalmente
 DEBORA SPENASSATO
Data: 12/02/2025 13:35:41-0300
Verifique em <https://validar.iti.gov.br>

Prof^a Doutora Débora Spenassato

(Avaliadora IMEF - FURG)
Documento assinado digitalmente
 FELIPE RICARDO SANTOS DE GUSMAO
Data: 12/02/2025 12:21:25-0300
Verifique em <https://validar.iti.gov.br>

**Prof^o Doutor Felipe Ricardo Santos
de Gusmão**

(Avaliador - IMEF - FURG)

Rio Grande, Rio Grande do Sul, Brasil
Janeiro, 2025

Agradecimentos

Agradeço ao Prof. Dr. Juliano Cesar Marangoni pela orientação, apoio e por embarcar comigo neste tema, contribuindo com sua dedicação e incentivo ao longo de todo o processo.

Minha gratidão à Profa. Dr. Raquel Nicolette, coorientadora, por ser a voz da razão e pelo suporte constante, cuja orientação foi fundamental para o desenvolvimento deste trabalho.

Expresso também meu reconhecimento aos professores Dr. Adilson da Silva Nunes e Dra. Juliana Ricardo da Silva Nunes, que me inspiraram a seguir na Matemática.

Agradeço, ainda, a todos os professores e colegas que cruzaram meu caminho durante o curso e contribuíram, de alguma forma, para minha formação.

Por fim, agradeço ao apoio imensurável da minha família.

“A Matemática é a Linguagem em que Deus escreveu o Universo.”
(Galileu Galilei)

Resumo

A análise multivariada, como área da estatística, destaca-se por suas amplas aplicações científicas, permitindo explorar múltiplas variáveis simultaneamente. O trabalho tem como objetivo aplicar técnicas da análise agrupamento para dados mistos (dados com variáveis quantitativas e qualitativas) no contexto das notificações de violência cujas as vítimas são exclusivamente do sexo feminino, utilizando o coeficiente de Gower e a distância Euclidiana como índices de dissimilaridade e o método de Ward como método de agrupamento. A partir dos dados do Sistema de Informação de Agravos de Notificação (SINAN), referentes ao ano de 2020, disponibilizados pelo DATASUS, buscando identificar grupos de vítimas entre os casos notificados no estado do Rio Grande do Sul, durante a pandemia de COVID-19. Assim, foi analisado o perfil das vítimas e as dinâmicas associadas à violência, a partir da aplicabilidade dos métodos estatísticos utilizados no estudo. Foram identificados três grupos principais de vítimas de violência no RS, referentes ao ano de 2020, utilizando as duas medidas de dissimilaridade diferentes.

Palavras-chaves: Violência; Mulher; Dados Mistos; Coeficiente de Gower; Método de Ward

Abstract

Multivariate analysis, as an area of statistics, stands out for its broad scientific applications, allowing the exploration of multiple variables simultaneously. The work aims to apply cluster analysis techniques to mixed data (data with quantitative and qualitative variables) in the context of reports of violence whose victims are exclusively female, using the Gower coefficient and the Euclidean distance as dissimilarity indices and the Ward method as a grouping method. Based on data from the Notifiable Diseases Information System (SINAN), referring to the year 2020, made available by DATASUS, seeking to identify groups of victims among the cases reported in the state of Rio Grande do Sul, during the COVID-19 pandemic. The study analyzed victim profiles and violence dynamics using applicable statistical methods. In 2020, three primary categories of violence victims in RS were identified using two different dissimilarity measures.

Keywords: violence; Women; Mixed Data; Gower's Coefficient; Ward's Method

Sumário

1	REFERENCIAL TEÓRICO	8
2	OBJETIVOS	10
2.1	Objetivo Geral	10
2.2	Objetivos Específicos	10
3	METODOLOGIA	11
3.1	Dados	11
3.2	Filtragem dos dados	12
3.2.1	Análise Estatística	15
3.3	Suporte Computacional	15
4	SUPORTE MATEMÁTICO	16
4.1	Análise Multivariada	16
4.2	Análise de Agrupamento	16
4.2.1	Métodos de Agrupamento	17
4.2.1.1	Métodos de Agrupamento Hierárquicos	17
4.2.1.2	Métodos de Agrupamento Não Hierárquicos	17
4.2.2	Primeira Etapa (Índices de Similaridade e Dissimilaridade)	18
4.2.3	Índice de Dissimilaridade de Gower	18
4.2.3.1	Exemplo Prático com Dados	19
4.2.4	Segunda Etapa: Método de agrupamento	20
4.2.5	Dendrograma	22
4.2.6	Análise Exploratória de Dados	23
5	RESULTADOS E DISCUSSÃO	24
5.1	Análise de agrupamento com a distância Euclidiana	25
5.2	Análise com Índice de Gower	26
5.3	Discussão	27
6	CONCLUSÃO	29
	REFERÊNCIAS	30
A	ANEXO - TABELAS DIVERSAS	32
A	ANEXO - FICHA VIOLÊNCIA	37

B	CODIGO	39
----------	---------------	-------	-----------

1 Referencial Teórico

A violência é um grave problema social e de saúde pública em várias partes do mundo, e no Brasil, não é diferente. De acordo com a Organização Mundial da Saúde (OMS), violência é definida como o uso intencional de força física ou poder, ameaçado ou real, contra si mesmo, outra pessoa, ou contra um grupo ou comunidade, que resulte ou tenha alta probabilidade de resultar em lesões, morte, danos psicológicos, subdesenvolvimento ou privação.

A institucionalização da abordagem da violência como problema de saúde ocorreu em 2001, com a Política Nacional de Redução da Morbimortalidade por Acidentes e Violências (BRASIL, 2001). A Declaração das Nações Unidas (ONU) sobre a Eliminação da Violência contra a Mulher define a violência contra a mulher como “qualquer ato de violência de gênero que causa, ou pode causar, dano físico, sexual ou mental ou sofrimento à mulher, incluindo a ameaça de tais atos, coerção ou privação arbitrária da liberdade, quer ocorra na vida pública ou privada”.

A violência de gênero é um fenômeno que atinge mulheres de diferentes classes sociais, idades e origens. Em 2024, a Secretaria Estadual da Saúde (SES) lançou o “Boletim Epidemiológico do Estado do Rio Grande do Sul: violência contra meninas e mulheres de 10 a 59 anos”. Os dados revelam que, entre 2018 e 2022, esse público feminino representou 78% das vítimas de violências notificadas no Rio Grande do Sul.

Bassanezi (2002) ressalta que a consistência e validação de uma teoria dependem de sua capacidade de ser interpretada e explicada por meio da linguagem matemática, destacando a importância central da matemática aplicada às ciências da saúde e sociais. Nestas ciências, a complexidade dos fenômenos estudados, como a violência e as dinâmicas epidemiológicas, exige o uso de diversas ferramentas para modelagem, análise e previsão.

No caso específico da violência contra a mulher, é imprescindível o uso da análise estatística para visualizar e analisar os dados disponíveis, que, além disso, pode fornecer um embasamento científico para a formulação de políticas públicas. A estatística, assim como a modelagem matemática, permite o estudo de fenômenos sociais, como a violência de gênero, com um grau de precisão que pode ser fundamental na criação de intervenções baseadas em evidências, visando mitigar esses problemas.

Globalmente, a pandemia agravou a violência doméstica, ao mesmo tempo em que reduziu o acesso a serviços de apoio às vítimas, como assistência social, saúde, segurança pública e justiça. Segundo Bueno (2021), desde os primeiros meses de isolamento social, a ONU Mulheres relatou, com base no aumento de pedidos de ajuda em linhas telefônicas de canais de atendimento, que houve um incremento de casos de violência doméstica em

todo o mundo e que as mulheres eram as principais vítimas.

A técnica mais comumente utilizada para identificar estruturas naturais entre as observações, com base em perfis multivariados, é a análise de agrupamento, que reúne objetos ou indivíduos em grupos, de forma que os objetos no mesmo grupo sejam mais parecidos entre si do que com os objetos de outros grupos. No contexto deste trabalho, essa abordagem permite identificar grupos de mulheres com perfis distintos, o que pode ajudar a identificar padrões e futuramente direcionar políticas públicas adequadas para diferentes realidades.

O método de Ward é particularmente eficaz para lidar com grandes conjuntos de dados, onde a similaridade entre os objetos é essencial para a formação de grupos homogêneos. Esse método minimiza a variância dentro dos grupos, garantindo que as observações em um mesmo agrupamento sejam mais semelhantes entre si do que em relação a outros grupos. A combinação do método de Ward com o coeficiente de Gower permite lidar com variáveis tanto qualitativas quanto quantitativas. Isso possibilita segmentar e analisar os diferentes perfis de mulheres vítimas de violência.

Assim, o objetivo deste estudo é analisar as notificações dos casos de violência contra a mulher através do ano de 2020 disponibilizadas pelo Sistema de Informação de Agravos de Notificação (SINAN).

2 Objetivos

2.1 Objetivo Geral

Identificar grupos no conjunto de notificações de violência registradas no estado do Rio Grande do Sul, com foco específico em vítimas do sexo feminino durante o primeiro ano (2020) da pandemia da covid-19.

2.2 Objetivos Específicos

- Identificar grupos homogêneos de casos de violência com base nas características das vítimas e seu relacionamento com seu agressor.
- Analisar agrupamento dos casos.
- Comparar os padrões de violência entre diferentes tipos de violência e características das vítimas.
- Contribuir para o fenômeno investigado a partir um modelo de análise multivariada de dados.

3 Metodologia

3.1 Dados

Os dados utilizados neste estudo referem-se às notificações de violência cujas as vítimas são exclusivamente do sexo feminino. Foram selecionados os tipos de violência física, psicológica/moral, financeira e/ou outros tipos, ocorridas no estado do Rio Grande do Sul, no período de 01/01/2020 a 31/12/2020, compreendendo o primeiro ano de pandemia da covid-19, notificados no DATASUS (Departamento de Informática do Sistema Único de Saúde). O conjunto de dados foi obtido através do Sistema de Informação de Agravos de Notificação (SINAN), o qual é uma das várias bases de dados geridas pelo DATASUS. O SINAN é alimentado, principalmente, pela notificação e investigação de casos de doenças e agravos que constam da lista nacional de doenças de notificação compulsória, mas é facultado a estados e municípios incluir outros problemas de saúde importantes em sua região, disponibilizados pelo DATASUS. A base de dados anonimizada está disponível para acesso público conforme a Lei de Acesso à Informação (Lei nº 12.527, de 18 de novembro de 2011), permitindo a democratização dos dados epidemiológicos no Brasil. Para acessar os dados, as seguintes instruções foram seguidas:

- Entre na página do DATASUS. [<http://datasus.saude.gov.br/>]
- Selecione a opção **ACESSO À INFORMAÇÃO** no menu superior;
- Entre em **SERVIÇOS** e clique em **TRANSFERÊNCIA/DOWNLOAD DE ARQUIVOS**;
- Escolha a base de dados do SINAN, modalidade **Dados**, e selecione o período (usando SHIFT ou CTRL para selecionar múltiplos anos);
- Escolha a área geográfica (selecione estados ou a sigla BR para dados nacionais);
- Clique em **ENVIAR**, e então faça o download dos arquivos selecionados no formato .zip.

Após o download, o arquivo compactado foi descompactado utilizando ferramentas como 7-ZIP ou WinZip. O manual, legislação, fichas de notificação e dicionários de dados associados à violência também foram acessados para referência. Na figura **I** pode ser visualizado a página de transferência de arquivos do DATASUS.

Figura 1 – Página de Transferência de Arquivos do DATASUS.

Transferência de Arquivos

Download de arquivos

The screenshot shows the 'Download de arquivos' interface with the following filters:

- Fonte:** A dropdown menu with 'SINAN - Sistema de Informações de Agravos de Notificação' selected.
- Modalidade:** A dropdown menu with 'Dados' selected.
- Tipo de Arquivo:** A dropdown menu with 'VIOL - Violência doméstica, sexual e/ou outras violências' selected.
- Ano:** A dropdown menu with '2021' selected.
- UF:** A dropdown menu with 'BR' selected.

3.2 Filtragem dos dados

Antes de realizar qualquer análise, é fundamental que os dados passem por um processo de filtragem. A filtragem de dados envolve a identificação e a correção de inconsistências, como valores ausentes, duplicados ou incorretos, além da padronização de formatos de dados, bem como, visando selecionar apenas as informações relevantes para o estudo, removendo que não sejam necessários para a análise em questão. Para manipular os arquivos no formato .dbc, foi utilizado o pacote read.dbc no software R, que permite a leitura dos dados em formatos proprietários do DATASUS.

Após a importação dos dados, foram realizadas as seguintes etapas de filtragem:

Filtro geográfico: Os dados foram filtrados para incluir apenas notificações do estado do Rio Grande do Sul;

Filtro de tipo de ocorrência: Focamos nas notificações de violência contra vítimas do sexo feminino, conforme os critérios definidos pela ficha de notificação do SINAN.

A base de dados inicial das análises consistiu em 16.046 observações de 40 variáveis selecionadas para análise conforme Tabela 1.

Tabela 1 – Quadro com os nomes dos campos das variáveis e suas descrições conforme o dicionário de dados do SINAN.

Nome do Campo	Descrição
Data da Notificação	Data de preenchimento da ficha de notificação.
Sexo	M- Masculino F- Feminino I- Ignorado
Raça/Cor	1 branca 2 preta 3 amarela 4 parda 5 indígena 9 Ignorado
Situação Conjugal	1 Solteiro 2 Casado/ União consensual 3 Viúvo 4 Separado 8. Não se aplica
Orientação Sexual	1 Heterossexual 2 Homossexual (gay/lésbica) 3 Bissexual 8 Não se aplica 9 Ignorado
Tipo de Violência - Física	1.Sim 2.Não 9. Ignorado
Tipo de Violência - Psicológica	1.Sim 2.Não 9. Ignorado
Tipo de Violência - Sexual	1.Sim 2.Não 9. Ignorado
Tipo de Violência - Financeira	1.Sim 2.Não 9. Ignorado
Tipo de Violência - Tortura	1.Sim 2.Não 9. Ignorado
Tipo de Violência - Negligência	1.Sim 2.Não 9. Ignorado
Tipo de Violência - Outros	1.Sim 2.Não 9. Ignorado
Número de Envolvidos	1 Um 2 Dois ou mais 9.Ignorado
Relação com a Pessoa Atendida -Pai	1.Sim 2.Não 9. Ignorado
Relação com a Pessoa Atendida- Mãe.	1.Sim 2.Não 9. Ignorado
Relação com a Pessoa Atendida - Padrasto	1.Sim 2.Não 9. Ignorado
Relação com a Pessoa Atendida Madrasta	1.Sim 2.Não 9. Ignorado
Relação com a Pessoa Atendida Conjuge	1.Sim 2.Não 9. Ignorado
Relação com a Pessoa Atendida ex-conjuge	1.Sim 2.Não 9. Ignorado
Relação com a Pessoa Atendida Namorado	1.Sim 2.Não 9. Ignorado
Relação com a Pessoa Atendida ex-namorado	1.Sim 2.Não 9. Ignorado
Relação com a Pessoa Atendida filho(a)	1.Sim 2.Não 9. Ignorado
Relação com a Pessoa Atendida irmão(ã)	1.Sim 2.Não 9. Ignorado
Relação com a Pessoa Atendida amigo(s)/conhecido(s)	1.Sim 2.Não 9. Ignorado
Relação com a Pessoa Atendida Desconhecido	1.Sim 2.Não 9. Ignorado
Relação com a Pessoa Atendida Cuidador	1.Sim 2.Não 9. Ignorado
Relação com a Pessoa Atendida Patrão/chefe	1.Sim 2.Não 9. Ignorado
Relação com a Pessoa Atendida Relação Institucional	1.Sim 2.Não 9. Ignorado
Relação com a Pessoa Atendida Policial/agente da lei	1.Sim 2.Não 9. Ignorado
Relação com a Pessoa Atendida própria pessoa	1.Sim 2.Não 9. Ignorado
Relação com a Pessoa Atendida outro tipo de relação	1.Sim 2.Não 9. Ignorado
Relação com a Pessoa Atendida Desconhecido	1.Sim 2.Não 9. Ignorado

Fonte: Elaborada pela autora com base nos dados do SINAN.

Após uma avaliação inicial das variáveis, o conjunto de dados passou por um processo de consolidação das mesmas. Pode-se citar os seguintes casos:

- Os tipos de violência originalmente registrados em múltiplas colunas (violência física, psicológica, sexual, financeira, negligência, entre outros) foram unificados em uma única variável qualitativa nominal;
- As relações entre vítima e agressor, previamente categorizadas em diversas descrições (pai, mãe, cônjuge, namorado, entre outras), foram unificados em uma única variável qualitativa nominal;
- As variáveis como raça, relação com parceiro e orientação sexual, foram recodificadas de categorias numéricas da ficha do SINAN para variáveis qualitativas nominais;
- Remoção de valores ausentes;
- Ponto de corte de idade: 18 anos ou mais.

Certos casos com informações classificadas como **não se aplica** e **ignorado** foram analisados conforme o dicionário de dados para decisão de exclusão.

Assim, ao final do processo de 16.046 observações extraídas do banco de dados previamente selecionado, foram mantidas 10.911, o que equivale a (67.99%) das observações extraídas, as mesmas podem ser visualizadas na Tabela 2.

De acordo com (MAGALHÃES; LIMA, 2012) note que as variáveis qualitativas ou categóricas representam atributos e/ou qualidades. Se tais variáveis têm uma ordenação natural, indicando intensidades crescentes de realização, então elas serão classificadas como qualitativas ordinais. Caso contrário, quando não é possível estabelecer uma ordem natural entre seus valores, elas são classificadas como qualitativas nominais. Variáveis como Sexo (feminino ou masculino) e Raça (branca, parda, indígena, amarela) são variáveis qualitativas nominais. Por outro lado, variáveis como Tamanho (pequeno, médio ou grande), Classe Social (baixa, média ou alta) são variáveis qualitativas ordinais. As variáveis quantitativas, variáveis de natureza numérica, podem ser subdivididas em discretas e contínuas. A grosso modo, variáveis quantitativas discretas podem ser vistas como resultantes de contagens, assumindo assim, em geral, valores inteiros. De uma maneira mais formal, o conjunto dos valores assumidos é finito ou enumerável. Já as variáveis quantitativas contínuas assumem valores em intervalos dos números reais e, geralmente, são provenientes de uma mensuração.

Tabela 2 – Quadro de descrição das Variáveis Analisadas

Nome do Campo	Descrição	Tipo de Variável
Idade	Idade \geq 18 anos	Quantitativa (Discreta)
Raça/Cor	Branca; Negra; Parda; Indígena; Amarela	Qualitativa (Nominal)
Situação Conjugal	Solteiro; Casado/União Consensual; Viúvo; Separado; Não se aplica	Qualitativa (Nominal)
Orientação Sexual	Heterossexual; Homossexual; Ignorado; Bissexual; Não se aplica; Não especificado	Qualitativa (Nominal)
Tipo de Violência	Física; Psicológica; Sexual; Financeira; Negligência; Outras	Qualitativa (Nominal)
Número de Envolvidos	Número de pessoas envolvidas na violência: 1 (Um); 2 (Mais de um)	Quantitativa (Discreta)
Relação com a Pessoa Atendida	Pai; Mãe; Madrasta; Cônjuge; Ex-cônjuge; Namorado; Filho(a); Irmão(ã); Amigo/Conhecido; Desconhecido; Cuidador; Patrão/Chefe; Relação institucional; Policial/Agente da lei; Própria pessoa; Outro	Qualitativa (Nominal)

Fonte: Elaborada pela autora com base nos dados do SINAN transformados.

3.2.1 Análise Estatística

Visando verificar a existência de grupos de vítimas mais semelhantes entre si, foi utilizada a análise de agrupamento hierárquico. O banco de dados sobre violência contra a mulher se caracterizou por possuir variáveis qualitativas e quantitativas, desta forma foi utilizado os índices de dissimilaridade distância euclidiana (JAMES et al., 2006) e Gower (GOWER, 1982) e o método de agrupamento Ward (WARD, 1963).

3.3 Suporte Computacional

Para a execução das análises estatísticas, foi utilizado o software R (R Core Team, 2025), por meio da interface RStudio (Posit Team, 2025). Inicialmente, foram empregadas bibliotecas como `read.dbc` (CARVALHO; ARSLAN, 2025) para a leitura dos dados, `dplyr` (WICKHAM et al., 2023) para a filtragem e manipulação do conjunto de dados, e `openxlsx` (WALKER; OEHLSCHLÄGEL; KRIJNEN, 2025) para a exportação e edição de arquivos Excel.

Para a análise de agrupamento hierárquico, foi utilizado o pacote `cluster` (MAECHLER et al., 2025), com o cálculo do índice de Gower como medida de dissimilaridade e o método de agrupamento de Ward. O pacote `dendroextras` (WHITE, 2025) foi empregado para a visualização do dendrograma, permitindo, em particular, o recorte dos grupos observados e a extração dos mesmos para análise. Os grupos formados foram exportados para uma planilha com o auxílio do pacote `openxlsx` (WALKER; OEHLSCHLÄGEL; KRIJNEN, 2025).

Por fim, o pacote `ggplot2` (WICKHAM, 2016) foi utilizado para a criação de gráficos que facilitaram a visualização e interpretação dos resultados.

O pacote `psych` foi utilizado para fazer a análise descritiva e exploratória, o pacote `autoEDA` foi utilizado para gráficos e o pacote `FactoMineR` a partir de componentes principais.

4 Suporte Matemático

4.1 Análise Multivariada

A análise multivariada é uma área da estatística que trata da análise de múltiplas variáveis em uma amostra, utilizando diferentes técnicas de acordo com o tipo de dado e os objetivos do estudo. Segundo Anderson (1984), as análises multivariadas podem ser classificadas em técnicas que permitem extrair informações sobre a interdependência entre as variáveis, como análise de agrupamento, análise fatorial e análise de componentes principais e em técnicas que exploram a dependência entre variáveis, como a análise de regressão multivariada, análise discriminante e análise de variância multivariada.

Análises multivariadas avaliam hipóteses cuja variável resposta é definida por múltiplas variáveis ao mesmo tempo, comumente expressa na forma de uma matriz quadrada ou de distância (dissimilaridade). Em geral, análises multivariadas têm três principais utilidades:

1. reduzir a dimensionalidade dos dados e encontrar a principal direção de variação dos mesmos;
2. testar relações entre matrizes, ou ainda;
3. encontrar diferenças entre grupos.

4.2 Análise de Agrupamento

A análise de agrupamento tem como objetivo identificar grupos naturais de indivíduos com base em medidas de similaridade ou dissimilaridade. A escolha adequada dessas medidas é crucial para a construção da análise (Zupan, 1982; Cormack, 1971). Nenhuma suposição é feita a priori a respeito do número de grupos ou de sua estrutura, sendo o agrupamento feito com base nas similaridades ou dissimilaridades. Muitas técnicas são propostas, pois não há, ainda, uma teoria generalizada e amplamente aceita (Johnson, 1982).

Medidas de dissimilaridade e similaridade desempenham um papel central nos algoritmos de agrupamento, ajudando a avaliar a distância entre duas observações. Neste trabalho, utilizamos técnicas de análise de agrupamento para identificar padrões entre vítimas de violência, construindo grupos mais similares ou diferentes entre si. Barroso e Artes (2003) destacam dois tipos de medidas de semelhança: medidas de similaridade,

em que um valor maior indica maior semelhança, e medidas de dissimilaridade, em que um valor maior indica menor semelhança.

Bussab et al. (1990) verificaram que a estrutura básica da aplicação das técnicas de análise de agrupamento pode ser decomposta nas seguintes etapas: definição de objetivos e critérios, escolha de variáveis e objetos, obtenção dos dados, tratamento das observações, escolha de critérios de similaridade ou dissimilaridade, adoção e execução de um algoritmo, apresentação dos resultados e avaliação dos mesmos. Muitas vezes, é necessário retornar a etapas anteriores para correções e aperfeiçoamentos.

4.2.1 Métodos de Agrupamento

4.2.1.1 Métodos de Agrupamento Hierárquicos

Segundo Zupan (1982), a análise de agrupamentos gera conexões hierárquicas ou ligações entre grupos de objetos ou indivíduos em diversas áreas do conhecimento. Métodos hierárquicos, como os métodos aglomerativo e divisivo, geram uma estrutura de árvore (dendograma) que organiza os indivíduos em sucessivos agrupamentos.

Os métodos hierárquicos baseiam-se na construção de uma matriz de dissimilaridade ou de similaridade, onde cada elemento descreve o grau de similaridade ou de dissimilaridade entre dois casos com base nas variáveis escolhidas. Entre os métodos hierárquicos aglomerativos, citamos o proposto por Ward (1963).

No método aglomerativo, inicialmente, cada elemento é considerado um grupo. Ao longo das etapas, esses elementos se agrupam até que, no final, exista um único grupo contendo todos os elementos. Já no método divisivo, o processo é inverso: todos os elementos começam em um único grupo e, ao longo das etapas, os grupos vão se dividindo até que cada grupo contenha apenas um elemento. A característica mais marcante dos métodos hierárquicos é que eles fornecem a trajetória de cada elemento até seu grupo final, gerando várias possibilidades de partições dos dados.

4.2.1.2 Métodos de Agrupamento Não Hierárquicos

Em contraste, os métodos não hierárquicos dividem os elementos em um número K de grupos definidos previamente, buscando otimizar a interpretação dos grupos (Bussab, 1990; Johnson, 1992). Esses métodos permitem que elementos sejam realocados entre os grupos durante o processo, tornando o método mais flexível. Um exemplo comum é o algoritmo k -means, que minimiza a variabilidade dentro dos grupos e maximiza a variabilidade entre eles.

4.2.2 Primeira Etapa (Índices de Similaridade e Dissimilaridade)

De maneira geral, as medidas de similaridade e de dissimilaridade são utilizadas quando trabalhamos com variáveis quantitativas e binárias (0 e 1). Quando lidamos com variáveis qualitativas, podemos seguir por dois caminhos:

- (a) transformando-as em variáveis quantitativas (variável numérica ordinal e/ou variável binária) e usando medidas de similaridade e/ou dissimilaridade
- (b) utilizando medidas de similaridade ou dissimilaridade adequadas para variável qualitativa.

Ambas as medidas de similaridade e dissimilaridade desempenham um papel central nos algoritmos de agrupamento, pois ajudam a avaliar se dois pontos estão próximos o suficiente para pertencerem ao mesmo grupo. Barroso e Artes (2003) destacam dois tipos de medidas de semelhança: medidas de similaridade, em que um valor maior indica maior semelhança, e medidas de dissimilaridade, em que um valor maior indica menor semelhança.

A distância Euclidiana (JAMES et al., 2006) é uma das medidas mais utilizadas para avaliar a dissimilaridade entre dois objetos ou observações. Ela é definida como:

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}, \quad (4.1)$$

onde d_{ij} representa a distância entre as observações i e j , e p é o número de variáveis consideradas. Nesse caso, a distância 0 (zero) indica que dois objetos ou observações são iguais.

Por outro lado, as medidas de similaridade quantificam a relação ou associação entre dois objetos. Um exemplo é a correlação de Pearson (PEARSON, 1895), definida como:

$$r_{ij} = \frac{\sum_{k=1}^p (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\sqrt{\sum_{k=1}^p (x_{ik} - \bar{x}_i)^2 \sum_{k=1}^p (x_{jk} - \bar{x}_j)^2}}, \quad (4.2)$$

onde r_{ij} representa o coeficiente de correlação entre as observações i e j , e \bar{x}_i e \bar{x}_j são as médias das variáveis para as respectivas observações. Um coeficiente de correlação $r = 0$ indica ausência de associação ou total diferença entre dois objetos ou observações.

4.2.3 Índice de Dissimilaridade de Gower

O índice de dissimilaridade de Gower é uma ferramenta extremamente útil para análises multivariadas em conjuntos de dados que incluem variáveis qualitativas e quanti-

tativas. Essa versatilidade permite sua aplicação em contextos que combinam diferentes tipos de informações, como idade e número de envolvidos (quantitativas), bem como raça e estado civil (qualitativas).

A distância de Gower mede a dissimilaridade entre indivíduos com base em diferentes tipos de variáveis, transformando-as em uma matriz de distâncias. Esse índice utiliza coeficientes binários para comparar variáveis e calcular a dissimilaridade entre os elementos, atribuindo uma pontuação geral que reflete a dissimilaridade.

De acordo com Gower (1971), a medida de proximidade deve estar definida no intervalo $[0, 1]$. O coeficiente é calculado considerando, para cada variável j , um coeficiente de similaridade s_j . A dissimilaridade entre dois elementos l e k é definida como:

$$d_{(l,k)} = \frac{\sum_{j=1}^p s_{jlk} w_j}{\sum_{j=1}^p w_j} \quad (4.3)$$

Onde:

- s_{jlk} : dissimilaridade entre os elementos l e k para a variável j , com valores no intervalo $[0, 1]$;
- w_j : peso atribuído à variável j .

Essa fórmula permite combinar diferentes tipos de variáveis ao calcular a dissimilaridade entre observações de maneira uniforme.

4.2.3.1 Exemplo Prático com Dados

Considere as seguintes cinco observações relacionadas a características de vítimas e agressores:

Idade	Tipo de violência	Relação com agressor	Raça	Orientação sexual	Situação conjugal	Número de envolvidos
23	Outras	Própria pessoa	Branca	Ignorado	Solteiro	1
50	Física	Amigo/Conhecido	Preta	Heterossexual	Solteiro	1
37	Física	Namorado	Branca	Heterossexual	Solteiro	1
57	Outras	Própria pessoa	Branca	Heterossexual	Casado	1
39	Física	Própria pessoa	Branca	Heterossexual	Casado	2

Tabela 3 – Dados transformados.

Para ilustrar o cálculo, considere as observações 1 e 2:

1. **Variáveis Quantitativas Discretas (Ex.: Idade):** O cálculo da dissimilaridade para a idade é realizado com base na diferença entre os valores observados, dividida pelo intervalo total da variável. Neste caso:

$$\text{Dissimilaridade}_{\text{Idade}} = \frac{|23 - 50|}{57 - 23} = \frac{27}{34} \approx 0,794.$$

2. **Variáveis Qualitativas (Ex.: Situação Conjugal):** Se as observações forem iguais, atribuímos dissimilaridade 0. Caso contrário, a dissimilaridade é 1. No caso da **situação conjugal** entre as observações 1 e 2, o valor é:

$$\text{Dissimilaridade}_{\text{Situação Conjugal}} = 0 \quad (\text{Ambos são Solteiro}).$$

3. **Cálculo Final da Dissimilaridade:** Somando as contribuições ponderadas de todas as variáveis (com pesos iguais $w_j = 1$), temos:

$$\text{Dissimilaridade Total} = \frac{\sum_j (\text{Dissimilaridade}_j \cdot w_j)}{\sum_j w_j}.$$

Após realizar os cálculos para todas as variáveis, obtemos uma matriz de dissimilaridade simétrica. Por exemplo, para as observações 1 e 2, o valor final de dissimilaridade é aproximadamente 0,684, considerando todas as variáveis.

- (a) A matriz de dissimilaridade gerada pelo índice de Gower é **simétrica**, ou seja, a distância entre as observações i e j é a mesma que a distância entre j e i ;
- (b) A diagonal principal da matriz apresenta valores iguais a zero, indicando que cada observação é idêntica a si mesma.

4.2.4 Segunda Etapa: Método de agrupamento

Os métodos de agrupamento, tanto hierárquicos quanto não hierárquicos, têm como base as distâncias calculadas entre os elementos, sejam elas dissimilaridades ou similaridades. Essas distâncias permitem que grupos de indivíduos mais similares sejam criados, enquanto indivíduos diferentes sejam alocados em grupos distintos.

No caso do método hierárquico de Ward, por exemplo, a escolha é baseada em minimizar a variabilidade intragrupo. Essa abordagem é particularmente útil para identificar padrões e gerar interpretações úteis em dados complexos, como aqueles relacionados às vítimas de violência.

Os métodos de análise de agrupamento estão relacionados à exploração de conjuntos de dados para avaliar se eles podem ser resumidos de forma significativa em um número relativamente pequeno de grupos de objetos ou indivíduos que se assemelham entre si e que são, em certos aspectos, diferentes dos indivíduos de outros grupos.

Em outras palavras, a análise de agrupamento busca descobrir agrupamentos naturais de indivíduos (ou variáveis) a partir dos dados observados, agrupando indivíduos com base na similaridade ou distâncias (dissimilaridades). A grosso modo, uma análise de agrupamento tenta resumir uma grande quantidade de dados e apresentá-la de maneira fácil de visualizar e entender (em geral, na forma de um dendrograma). No entanto, os

resultados da análise podem não refletir necessariamente toda a informação originalmente contida na matriz de dados.

Uma grande variedade de métodos de agrupamento foi desenvolvida ao longo das últimas quatro décadas. O método de Ward, apresentado em um artigo no *Journal of the American Statistical Association* (1963), é um método de agrupamento hierárquico. Ward apresenta uma abordagem para o problema de agrupamento (clustering) com foco em organizar grandes conjuntos de objetos, símbolos ou pessoas em grupos menores e mutuamente exclusivos, de forma que os membros de cada grupo sejam o mais semelhantes possível. Essa organização facilita a análise e a compreensão de relações dentro de grandes conjuntos e melhora a eficiência na gestão das informações.

Ward (1963) destaca que, embora o agrupamento frequentemente resulte em perda de informação, é possível quantificar essa perda por meio de um **número que reflete o valor**. Diferente de formulações anteriores, como as de Cox e Fisher ((COX, 1957))((FISHER, 1958)), que lidavam com variáveis individuais ou minimização de somas de quadrados sobre uma única variável, Ward introduz um método que considera múltiplas variáveis.

O método de Ward é baseado no critério de quadrados mínimos (*OLS - Ordinary Least Squares*), o mesmo utilizado para ajustar um modelo linear. O objetivo é definir os grupos de maneira que a soma de quadrados (i.e. similar ao erro quadrado da ANOVA) dentro dos grupos seja minimizada (Borcard, Gillet, e Legendre 2018). O método de Ward busca formar partições onde os grupos sejam o mais homogêneos possíveis internamente e o mais heterogêneos possíveis entre si.

O método começa tratando cada item como um grupo e, em seguida, agrupa-os de acordo com a menor distância $\min d_{jk}$. Para um grupo i , define-se o ESS_i , que é a soma dos desvios de cada item em relação à média do grupo:

$$ESS_i = \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i) \quad (4.4)$$

onde n_i é o número de elementos no grupo i .

No passo k , a soma dos quadrados dentro dos grupos é:

$$\sum_{i=1}^k SSR_i = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i) \quad (4.5)$$

A distância entre os grupos C_l e C_i é definida como:

$$d(C_l, C_i) = \frac{n_l n_i}{n_l + n_i} (\bar{X}_l - \bar{X}_i) (X_{ij} - \bar{X}_i) \quad (4.6)$$

onde $d(C_l, C_i)$ é a soma dos quadrados entre os grupos C_l e C_i .

Em cada passo, dois grupos são combinados de acordo com a menor distância $\min d(\cdot)$. A distância $\min d(\cdot)$ é a diferença entre o valor de SSR antes e depois da combinação dos grupos, resultando no menor valor de SSR . O método de Ward leva em consideração a diferença de tamanho dos grupos ao realizar essa combinação.

4.2.5 Dendrograma

Nos métodos hierárquicos, o dendrograma é uma representação gráfica e hierárquica do agrupamento dos elementos analisados. Cada folha do dendrograma corresponde a uma observação. Conforme subimos na árvore, observações similares entre si são combinadas em ramos, que, por sua vez, são agrupados em alturas maiores.

A altura da fusão, indicada no eixo vertical, reflete o nível de dissimilaridade entre as observações. Quanto maior a altura, menos similares as observações são. É importante destacar que as conclusões sobre a proximidade de duas observações devem ser baseadas exclusivamente na altura em que os ramos contendo essas observações são unidos pela primeira vez. A proximidade ao longo do eixo horizontal não deve ser interpretada como um critério de similaridade.

Figura 2 – Exemplo de dendrograma.

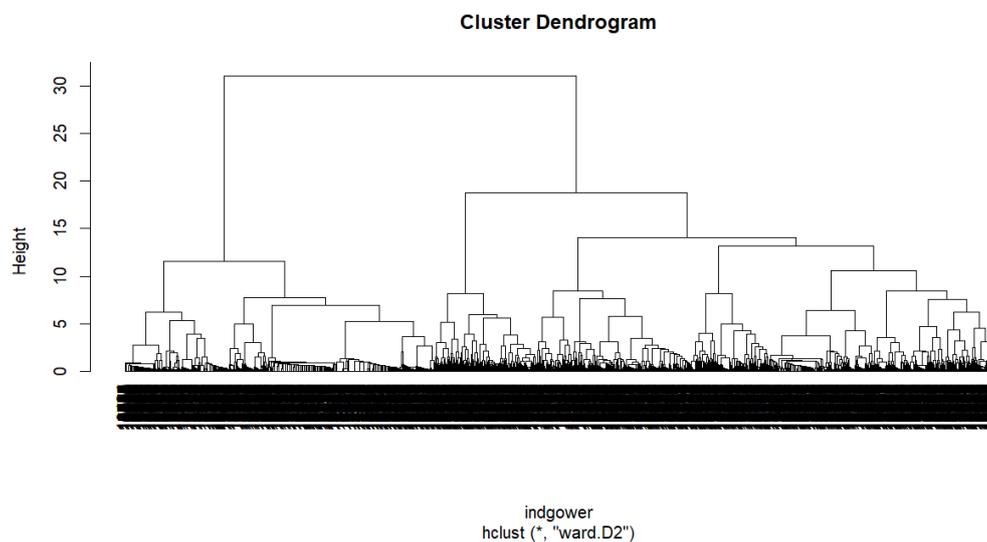
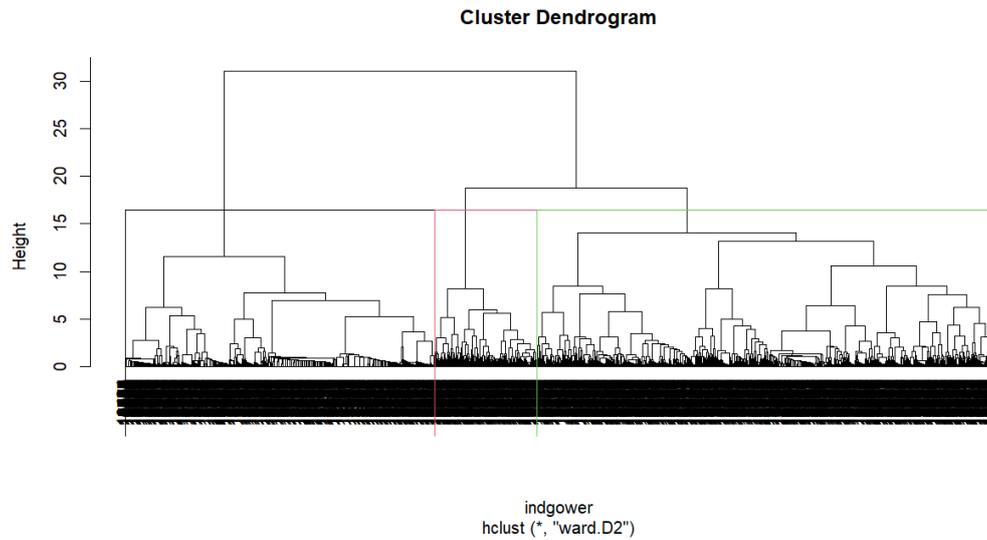


Figura 3 – Corte Dendrograma.



Conforme observado no dendrograma exibido na Figura 2, cada observação individual é conectada a outra observação através dos ramos com base no grau de similaridade entre elas. Na Figura 3, destaca-se a divisão em três grupos, resultado da escolha de um ponto de corte específico na árvore.

A escolha da altura do corte no dendrograma foi determinada por forma supervisionada (determinada pela autora) a partir de uma interpretação visual, considerando os degraus evidentes. Esse método baseia-se na identificação de diferenças perceptíveis entre as distâncias das observações, onde uma maior distância indica maior dissimilaridade entre os grupos. Essa abordagem permitiu identificar três grupos principais nos dados analisados, como mostrado na Figura 3.

4.2.6 Análise Exploratória de Dados

Inicialmente, foi realizada uma análise exploratória dos dados para compreender melhor o comportamento das variáveis e a estrutura do agrupamento. Essa etapa é fundamental para identificar possíveis variáveis que possam dificultar o agrupamento ou influenciar negativamente os resultados.

A análise exploratória tem como objetivo *brincar* com o conjunto de dados, permitindo uma compreensão mais intuitiva e visual das informações. Para isso, foram calculadas medidas descritivas como média, moda, mediana, variância, desvio padrão e frequência relativa. Essas estatísticas principais permitiram uma visão geral dos dados, identificando tendências e outliers, que podem ser importantes na etapa subsequente de análise multivariada.

5 Resultados e Discussão

A primeira parte da análise consiste em uma análise exploratória geral das variáveis. A Tabela 4 apresenta o resumo das variáveis quantitativas.

Tabela 4 – Resumo das Variáveis quantitativas

	Idade	Nº Agressores
Min.	18	1
1ºQ	25	1
Mediana	34	1
Média	37,17	1,12
3ºQ	46	1
Max	102	2

As Tabelas 5 e 6 apresentam os resumos gerais das variáveis qualitativas.

Tabela 5 – Resumo das variáveis qualitativas

tipo_violência	relacao_agressor	raca
Física :4102	Própria Pessoa :3894	Amarela : 42
Outras :3896	Cônjuge :2334	Branca :8635
Psicológica :1178	Ex-cônjuge :1204	Ignorado: 192
Física, Psicológica: 528	Outra Relação : 550	Indígena: 83
Sexual : 502	Amigo/Conhecido: 503	Parda :1120
Negligência : 197	Desconhecido : 460	Preta : 839
(Other) : 508	(Other) :1966	

Algumas dessas variáveis apresentaram muitas categorias que não foram apresentadas nas tabelas anteriores. Um exemplo é a variável *tipo_violencia*, a qual apresentou 52 subdivisões que podem ser encontradas no Anexo A.

A variável *relacao_agressor* apresenta 98 subdivisões.

Tabela 6 – Resumo das variáveis qualitativas

orientacao_sexual	situacao_conjugal
Bissexual : 39	Casado/União consensual:3896
Heterossexual:7552	Não especificado :1334
Homossexual : 193	Não se aplica : 134
Ignorado :2364	Separado : 995
Não se aplica: 763	Solteiro :4189
	Viúvo : 363

O próximo passo da análise foi a realização da Análise de Agrupamento Hierárquica das mulheres vítimas de violência no RS, referentes ao ano de 2020. Através dos diferentes índices de dissimilaridade, distância Euclidiana e coeficiente de Gower.

5.1 Análise de agrupamento com a distância Euclidiana

A análise de agrupamento hierárquica, utilizando a distância Euclidiana, revelou três grupos distintos das 10.911 vítimas de violência. Os resultados são apresentados a seguir, com destaque para os perfis identificados em cada grupo. Na Figura 4, pode-se observar o dendrograma cujo corte dos grupos foi feito através de um procedimento supervisionado.

Figura 4 – Dendrograma.

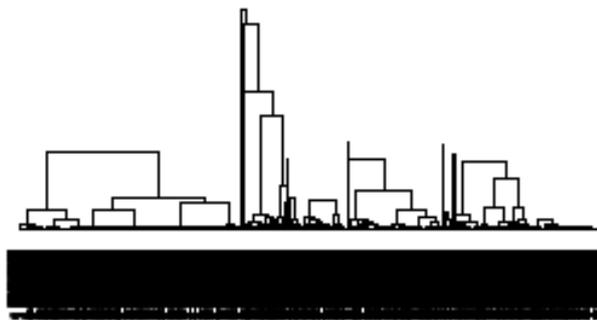
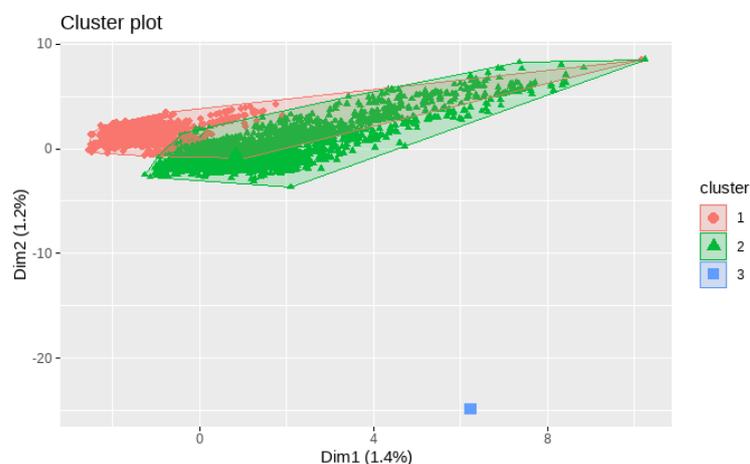


Figura 5 – Grupos distância Euclidiana



O grupo 1, identificado por 1 na Figura 5, compreendeu 36,25% de todas as observações. Destacaram-se vítimas com menos de 20 anos e a violência autoprovocada.

A maioria dos agressores, mais de 95%, foi a própria pessoa, enquanto as demais relações apresentavam percentual inferior a 5%. A maior parte das vítimas desse grupo era solteira, heterossexual e branca.

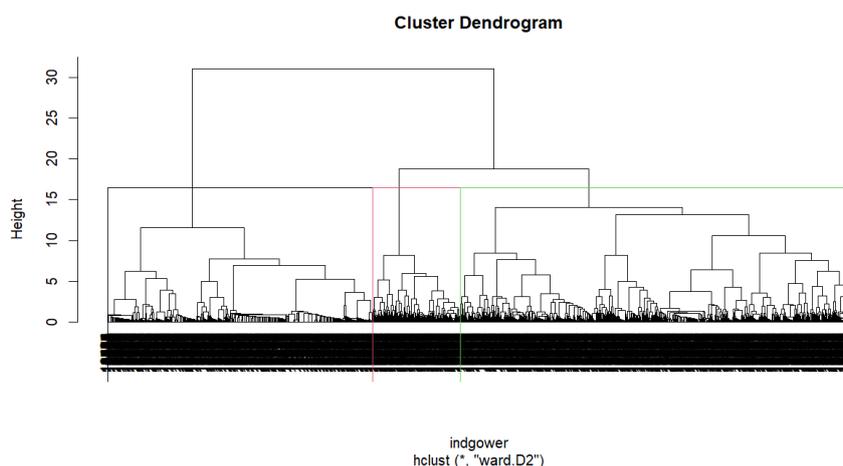
O grupo 2, identificado por 2 na Figura 5 compreendeu 63,74% das observações. Destacou-se a violência física, ocorrendo em mais de 50% dos casos.. Havia um equilíbrio entre vítimas solteiras e casadas (37,41% e 37,96%, respectivamente). O agressor mais comum foi o cônjuge, seguido pelo ex-cônjuge.

O grupo 3, identificado por 3 na Figura 5, foi composto por uma vítima de 32 anos, que sofreu mais de uma forma de violência, incluindo psicológica, financeira, tortura e negligência. Os agressores envolvidos foi tanto o cônjuge quanto um agente da lei. A vítima era parda e casada.

5.2 Análise com Índice de Gower

A análise de agrupamento utilizando o coeficiente de Gower também revelou três grupos principais. O dendrograma correspondente é apresentado na Figura 6. O corte do dendrograma para determinação do número de grupos foi feito através de um procedimento supervisionado.

Figura 6 – Dendrograma.



No grupo 1, destacou-se a faixa etária das vítimas (predominantemente dos 18 a 37 anos), compreendendo aproximadamente 61,2% das vítimas do grupo e o tipo de violência autoprovocada, como a tentativa de suicídio e a automedicação, compreendendo aproximadamente 98,3% das vítimas do grupo. Em 100% dos casos, apenas um indivíduo esteve envolvido como agressor. Em relação à raça, 83,12% das vítimas tinham esse dado registrado. Sobre a orientação sexual, 62,35% das vítimas eram heterossexuais. Quanto à situação conjugal, 38,93% eram solteiras e 32,98% casadas.

O grupo 2, compreendeu 1.281 de 10.928 observações (11,17%). Destacou-se o envolvimento de dois ou mais agressores (95,9%), sendo as principais formas de violência a física (59,6%) e a físico-psicológica (11,16%). Os principais agressores foram o cônjuge (19,1%) e o amigo/conhecido (13,8%). Esse grupo apresentou a maior variação nas relações com o agressor, pois, em menos de 5% dos casos, havia apenas um agressor.

O grupo 3, compreendeu 52,64% das observações. Destacou-se o envolvimento de um único agressor (98% dos casos). As formas mais frequentes de violência foram a física (57,9%) e a psicológica (18,8%).

Comparando os dois métodos, os dois métodos captaram 3 grupos distintos, como determinado. Entretanto, constatou-se uma diferença no número de observações em cada grupo. Para o grupo 1 da distância Euclidiana possuía 3.956 observações, passando então para 3750 com Gower (redução de aproximadamente de 5% das observações). Para o grupo 2 da distância Euclidiana apresentou 6.954 observações, passando então para 6.311 com Gower (redução de aproximadamente 9% das observações). O grupo 3, com a distância Euclidiana possuía uma única observação, passando então para 811 observações com o índice de Gower.

5.3 Discussão

A diferença constatada entre os dois índices pode estar associada à metodologia utilizada na formação dos grupos, todos definidos de maneira supervisionada. Os agrupamentos foram realizados considerando diferentes critérios de dissimilaridade, o que pode ter influenciado a composição e o número de observações em cada grupo. No caso do método baseado na distância de Gower, a consideração simultânea de variáveis qualitativas e quantitativas preservou melhor as relações entre as observações, refletindo uma segmentação mais equilibrada. Já a abordagem utilizando a distância Euclidiana, que exige a conversão de variáveis qualitativas em binárias e posteriormente quantitativas, pode ter levado a uma distribuição diferenciada dos grupos devido ao impacto da normalização aplicada nesse processo.

A formação dos grupos pode ter sido influenciada pela transformação das variáveis qualitativas, possivelmente resultando em padrões distintos nos agrupamentos gerados por cada método. A análise dessas diferenças e suas implicações pode ser implementada em estudos futuros, incluindo simulações com dados mistos e comparações com métodos não supervisionados para avaliar as abordagens empregadas.

Esses dados permitem delinear um perfil preliminar das vítimas desse grupo, contribuindo para análises mais direcionadas no combate à violência. Segundo dados do Instituto Brasileiro de Geografia e Estatística (IBGE), obtidos por meio da plataforma SIDRA com filtro para a população a partir de 18 anos, o Rio Grande do Sul registrou, em 2022, um total de 13.358.025 habitantes que se autodeclararam brancos, dos quais 53,04% eram mulheres. Além disso, o estado contabilizou 2.341.181 pessoas pardas (49,75% mulheres), 1.125.725 pessoas pretas (51,81% mulheres), 44.520 indígenas (50,80% mulheres) e 13.785 pessoas amarelas (52,29% mulheres). O estado se destaca por apresentar a maior proporção de pessoas brancas entre todas as unidades federativas do Brasil (IBGE, 2022).

Em relação à orientação sexual, 62,35% das vítimas identificaram-se como heterossexuais. Essa composição demográfica não se mostra tão relevante ao analisar os dados sobre violência, uma vez que a predominância de determinados grupos étnicos e orientação sexual pode influenciar os padrões de registro e notificação de casos, de forma que a predominância na região estabelecida, neste caso no Rio Grande do Sul, não seja uma variável de tão importância quanto as outras na formação dos grupos.

Os dados nacionais indicam que, durante a pandemia da covid-19, a ideação suicida cresceu significativamente no Brasil, afetando mais mulheres. No primeiro semestre de 2020, mais de 70% das pessoas que passaram a apresentar ideação suicida eram do sexo feminino (USP, 2020). Esses números reforçam a necessidade de analisar os impactos sociais e psicológicos da pandemia, especialmente para o público feminino, o que diz respeito ao dados observados.

6 Conclusão

Este trabalho teve como objetivo identificar e analisar grupos de vítimas de violência no Rio Grande do Sul durante o primeiro ano da pandemia de covid-19, com foco específico em mulheres. Para isso, foram utilizadas técnicas de análise de agrupamento, aplicando os métodos da distância Euclidiana e do coeficiente de Gower, permitindo segmentar as vítimas com base em múltiplas variáveis.

A partir das análises, foi possível identificar perfis distintos de vítimas e padrões de violência, contribuindo para uma compreensão inicial do fenômeno. Além disso, a comparação entre os métodos evidenciou como diferentes critérios de dissimilaridade influenciam a formação dos grupos, ressaltando a importância da escolha metodológica na análise de dados complexos.

Os resultados deste estudo reforçam a relevância da aplicação de técnicas multivariadas para investigar a violência de gênero, demonstrando que a análise de agrupamento pode ser uma ferramenta útil para compreender padrões do fenômeno. As informações obtidas podem auxiliar no desenvolvimento de políticas públicas mais direcionadas e eficazes, apoiando ações de prevenção e suporte às vítimas.

Para estudos futuros, recomenda-se a ampliação do escopo da análise, incorporando dados de outros períodos para avaliar mudanças nos padrões de violência ao longo do tempo. Além disso, a inclusão de mais variáveis disponíveis no banco de dados, utilizando métodos de interpolação para minimizar a perda de informações devido à filtragem, pode enriquecer a análise e fornecer um panorama ainda mais detalhado.

Dessa forma, este trabalho não apenas contribui para a análise estatística da violência contra mulheres no Rio Grande do Sul, mas também abre caminho para novas investigações e para o aprimoramento das políticas de enfrentamento dessa grave questão social.

Referências

- BARROSO, L. P.; ARTES, R. *Análise Multivariada*. Lavras: UFLA, 2003. 157 p. Nenhuma citação no texto.
- BRASIL, M. d. S. *Política Nacional de Redução de Morbimortalidade por Acidentes e Violência (PNRMAV)*. 2001. Portaria GM/MS N^o 737 de 16 de maio de 2001, Diário Oficial da União, 17 maio 2001. Citado na página [8](#).
- CARVALHO, W. M. de; ARSLAN, Y. E. D. *read.dbc: Read DBC Files*. [S.l.], 2025. R package version 0.3.2. Disponível em: <https://CRAN.R-project.org/package=read.dbc>. Citado na página [15](#).
- COX, D. R. Note on grouping. *Journal of the American Statistical Association*, v. 52, p. 543–547, 1957. Citado na página [21](#).
- FISHER, W. D. On grouping for maximum homogeneity. *Journal of the American Statistical Association*, v. 53, p. 789–798, 1958. Citado na página [21](#).
- GOWER, J. C. Euclidean distance geometry. *Mathematical Scientist*, v. 7, n. 1, p. 1–14, 1982. Citado na página [15](#).
- JAMES, G. et al. *An Introduction to Statistical Learning: With Applications in R*. [S.l.]: Springer, 2006. Citado 2 vezes nas páginas [15](#) e [18](#).
- JUNIOR, A. A. B. *Análise Multivariada: Aula 5 - Análise de Agrupamentos (Clusters)*. 2016. Acesso em: 19 set. 2024. Disponível em: https://www2.ufjf.br/lates//files/2016/12/Conte%C3%BAdo-5-%E2%80%93-A_An%C3%A1lise-de-cluster-AA.pdf. Nenhuma citação no texto.
- MAECHLER, M. et al. *cluster: Cluster Analysis Basics and Extensions*. [S.l.], 2025. R package version 2.1.5. Disponível em: <https://CRAN.R-project.org/package=cluster>. Citado na página [15](#).
- MAGALHÃES, M. N.; LIMA, A. C. P. de. *Noções de Probabilidade e Estatística*. 2. ed. São Paulo: Edgard Blücher, 2012. ISBN 9788521206419. Citado na página [14](#).
- NATIONS, U. *Declaration on the elimination of violence against women*. New York: [s.n.], 1993. http://www.un.org/en/ga/search/view_doc.asp?symbol=A/RES/48/104. Resolution A/RES/48/104, accessed March 11, 2015. Nenhuma citação no texto.
- OLIVEIRA, N. L. d. *Análise de Agrupamento Hierárquicos*. Trabalho de Conclusão de Curso (Graduação em Estatística) — Universidade Federal do Ceará, Fortaleza, 2022. Nenhuma citação no texto.
- PEARSON, K. Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, v. 58, p. 240–242, 1895. Citado na página [18](#).
- Posit Team. *RStudio: Integrated Development for R*. Boston, MA, 2025. Versão 2023.12. Disponível em: <https://posit.co/>. Citado na página [15](#).

R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria, 2025. Versão 4.3.2. Disponível em: [<https://www.R-project.org/>](https://www.R-project.org/). Citado na página [15](#).

USP, J. da. *Ideação suicida cresceu na pandemia: mulheres e quem passou por situações adversas na infância foram mais afetados*. 2020. Acessado em: 28 jan. 2025. Disponível em: [ideacao-suicida-cresceu-na-pandemia-mulheres-e-quem-passou-por-situacoes-adversas-na-infancia-fo](https://jornal.usp.br/ciencias/ideacao-suicida-cresceu-na-pandemia-mulheres-e-quem-passou-por-situacoes-adversas-na-infancia-fo)[>](#). Citado na página [28](#).

WALKER, A.; OEHLISCHLÄGEL, P. L.; KRIJNEN, M. *openxlsx: Read, Write and Edit xlsx Files*. [S.l.], 2025. R package version 4.2.6.1. Disponível em: <https://CRAN.R-project.org/package=openxlsx>. Citado na página [15](#).

WARD, J. J. H. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, v. 58, n. 301, p. 236–244, Mar 1963. Citado na página [15](#).

WHITE, M. *dendroextras: Tree Extensions for 'dendrogram' Objects*. [S.l.], 2025. R package version 1.2.0. Disponível em: <https://CRAN.R-project.org/package=dendroextras>. Citado na página [15](#).

WICKHAM, H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. Disponível em: <https://ggplot2.tidyverse.org/>. Citado na página [15](#).

WICKHAM, H. et al. *dplyr: A Grammar of Data Manipulation*. [S.l.], 2023. R package version 1.1.3. Disponível em: <https://CRAN.R-project.org/package=dplyr>. Citado na página [15](#).

A Anexo - Tabelas diversas

Tabela 7 – Tipos de violência

	tipo_violencia
Financeira	41
Financeira, Negligência	1
Física	4102
Física, Financeira	3
Física, Financeira, Outras	1
Física, Negligência	1
Física, Outras	30
Física, Psicológica	528
Física, Psicológica, Financeira	29
Física, Psicológica, Financeira, Negligência	2
Física, Psicológica, Financeira, Outras	5
Física, Psicológica, Financeira, Tortura	4
Física, Psicológica, Financeira, Tortura, Negligência	1
Física, Psicológica, Financeira, Tortura, Negligência, Outras	1
Física, Psicológica, Negligência	5
Física, Psicológica, Outras	18
Física, Psicológica, Sexual	26
Física, Psicológica, Sexual, Financeira	3
Física, Psicológica, Sexual, Financeira, Negligência	1
Física, Psicológica, Sexual, Financeira, Tortura	1
Física, Psicológica, Sexual, Financeira, Tortura, Negligência	1
Física, Psicológica, Sexual, Outras	1
Física, Psicológica, Sexual, Tortura	6
Física, Psicológica, Tortura	45
Física, Psicológica, Tortura, Outras	3
Física, Sexual	24
Física, Sexual, Negligência	1
Física, Sexual, Tortura	3
Física, Tortura	6
Não especificado	122
Negligência	197
Outras	3896

Psicológica	1178
Psicológica, Financeira	19
Psicológica, Financeira, Negligência	5
Psicológica, Financeira, Outras	2
Psicológica, Financeira, Tortura	6
Psicológica, Financeira, Tortura, Negligência	1
Psicológica, Financeira, Tortura, Outras	5
Psicológica, Negligência	2
Psicológica, Negligência, Outras	1
Psicológica, Outras	42
Psicológica, Sexual	10
Psicológica, Sexual, Financeira	1
Psicológica, Tortura	13
Psicológica, Tortura, Outras	3
Sexual	502
Sexual, Financeira	2
Sexual, Negligência	1
Sexual, Outras	1
Tortura	9

	relacao_agressor
Amigo/Conhecido	503
Amigo/Conhecido, Desconhecido	11
Amigo/Conhecido, Desconhecido, Relação Institucional	1
Amigo/Conhecido, Outra Relação	8
Amigo/Conhecido, Patrão/Chefe	1
Amigo/Conhecido, Própria Pessoa	3
Cônjuge	2334
Cônjuge, Amigo/Conhecido	9
Cônjuge, Cuidador	1
Cônjuge, Desconhecido	2
Cônjuge, Ex-cônjuge	4
Cônjuge, Ex-namorado	1
Cônjuge, Filho(a)	19
Cônjuge, Filho(a), Amigo/Conhecido, Policial/Agente da Lei	1
Cônjuge, Irmão(ã)	4
Cônjuge, Namorado	3
Cônjuge, Namorado, Ex-namorado	1
Cônjuge, Namorado, Ex-namorado, Filho(a), Irmão(ã)	1
Cônjuge, Outra Relação	23
Cônjuge, Policial/Agente da Lei	1
Cônjuge, Própria Pessoa	6
Cuidador	20
Cuidador, Outra Relação	3
Cuidador, Relação Institucional	1
Desconhecido	460
Desconhecido, Cuidador	1
Desconhecido, Outra Relação	2
Desconhecido, Patrão/Chefe	3
Desconhecido, Própria Pessoa	1

	Ex-cônjuge	1204
	Ex-cônjuge, Amigo/Conhecido	5
Ex-cônjuge, Amigo/Conhecido, Desconhecido		1
	Ex-cônjuge, Desconhecido	2
	Ex-cônjuge, Ex-namorado	1
	Ex-cônjuge, Filho(a)	7
Ex-cônjuge, Filho(a), Outra Relação		1
	Ex-cônjuge, Outra Relação	13
Ex-cônjuge, Policial/Agente da Lei		1
	Ex-cônjuge, Própria Pessoa	3
	Ex-namorado	258
	Ex-namorado, Amigo/Conhecido	2
	Ex-namorado, Desconhecido	1
Ex-namorado, Filho(a), Irmão(ã), Amigo/Conhecido, Desconhecido		1
	Ex-namorado, Outra Relação	3
	Ex-namorado, Patrão/Chefe	1
	Filho(a)	452
	Filho(a), Amigo/Conhecido	2
	Filho(a), Cuidador	6
	Filho(a), Desconhecido	1
	Filho(a), Irmão(ã)	7
Filho(a), Irmão(ã), Outra Relação		1
	Filho(a), Outra Relação	32
Filho(a), Policial/Agente da Lei		1
	Filho(a), Própria Pessoa	3
Filho(a), Relação Institucional		1
	Irmão(ã)	219
	Irmão(ã), Amigo/Conhecido	2
	Irmão(ã), Outra Relação	13
	Irmão(ã), Própria Pessoa	1
	Madrasta	3
Madrasta, Amigo/Conhecido, Outra Relação		1
	Madrasta, Outra Relação	1
	Mãe	31
	Mãe, Cônjuge	1
	Mãe, Filho(a)	1
	Mãe, Irmão(ã)	6
	Mãe, Outra Relação	7
	Mãe, Padrasto	3
	Mãe, Padrasto, Irmão(ã)	1
	Mãe, Própria Pessoa	3
	Namorado	360
	Namorado, Amigo/Conhecido	3
Namorado, Amigo/Conhecido, Outra Relação		1
	Namorado, Filho(a)	1
	Namorado, Outra Relação	1
	Não especificado	152
	Outra Relação	550

Padrasto	43
Padrasto, Amigo/Conhecido	2
Padrasto, Cônjuge	1
Padrasto, Ex-cônjuge	1
Padrasto, Irmão(ã)	1
Padrasto, Madrasta	1
Padrasto, Outra Relação	1
Pai	69
Pai, Cônjuge	3
Pai, Cônjuge, Outra Relação	1
Pai, Desconhecido	2
Pai, Filho(a)	3
Pai, Irmão(ã)	3
Pai, Madrasta	1
Pai, Mãe	6
Pai, Mãe, Irmão(ã)	1
Pai, Mãe, Outra Relação	1
Pai, Padrasto	1
Pai, Padrasto, Cônjuge	1
Patrão/Chefe	26
Patrão/Chefe, Outra Relação	1
Patrão/Chefe, Relação Institucional	1
Policial/Agente da Lei	17
Policial/Agente da Lei, Outra Relação	1
Policial/Agente da Lei, Própria Pessoa	1
Própria Pessoa	3894
Própria Pessoa, Outra Relação	5
Relação Institucional	23
Relação Institucional, Própria Pessoa	1

A Anexo - Ficha Violência

República Federativa do Brasil Ministério da Saúde		SINAN SISTEMA DE INFORMAÇÃO DE AGRAVOS DE NOTIFICAÇÃO FICHA DE NOTIFICAÇÃO INDIVIDUAL		Nº	
Caso suspeito ou confirmado de violência doméstica/intrafamiliar, sexual, autoprovocada, tráfico de pessoas, trabalho escravo, trabalho infantil, tortura, intervenção legal e violências homofóbicas contra mulheres e homens em todas as idades. No caso de violência extrafamiliar/comunitária, somente serão objetos de notificação as violências contra crianças, adolescentes, mulheres, pessoas idosas, pessoa com deficiência, indígenas e população LGBT.					
Dados Gerais	1	Tipo de Notificação		2 - Individual	
	2	Agravado/doença	VIOLÊNCIA INTERPESSOAL/AUTOPROVOCADA	Código (CID10) Y09	
	3	Data da notificação			
	4	UF	5	Município de notificação	Código (IBGE)
	6	Unidade Notificadora	<input type="checkbox"/> 1- Unidade de Saúde <input type="checkbox"/> 2- Unidade de Assistência Social <input type="checkbox"/> 3- Estabelecimento de Ensino <input type="checkbox"/> 4- Conselho Tutelar <input type="checkbox"/> 5- Unidade de Saúde Indígena <input type="checkbox"/> 6- Centro Especializado de Atendimento à Mulher <input type="checkbox"/> 7- Outros		
	7	Nome da Unidade Notificadora	Código Unidade	9	
8	Unidade de Saúde	Código (CNES)	Data da ocorrência da violência		
Notificação Individual	10			Nome do paciente	
	11			Data de nascimento	
	12	(ou) Idade	<input type="checkbox"/> 1- Hora <input type="checkbox"/> 2- Dia <input type="checkbox"/> 3- Mês <input type="checkbox"/> 4- Ano	13	Sexo
			<input type="checkbox"/> M - Masculino <input type="checkbox"/> F - Feminino <input type="checkbox"/> I - Ignorado	14	Gestante
			<input type="checkbox"/> 1-1º Trimestre <input type="checkbox"/> 2-2º Trimestre <input type="checkbox"/> 3-3º Trimestre <input type="checkbox"/> 4- Idade gestacional ignorada <input type="checkbox"/> 5- Não <input type="checkbox"/> 6- Não se aplica	15	
			<input type="checkbox"/> 1- Branca <input type="checkbox"/> 2- Preta <input type="checkbox"/> 3- Amarela <input type="checkbox"/> 4- Parda <input type="checkbox"/> 5- Indígena <input type="checkbox"/> 9- Ignorado		
16		Escolaridade			
0- Analfabeto 1- 1ª a 4ª série incompleta do EF (antigo primário ou 1º grau) 2- 4ª série completa do EF (antigo primário ou 1º grau) 3- 5ª a 8ª série incompleta do EF (antigo ginásio ou 1º grau) 4- Ensino fundamental completo (antigo ginásio ou 1º grau) 5- Ensino médio incompleto (antigo colegial ou 2º grau) 6- Ensino médio completo (antigo colegial ou 2º grau) 7- Educação superior incompleta 8- Educação superior completa 9- Ignorado 10- Não se aplica					
17		Número do Cartão SUS			
18		Nome da mãe			
Dados de Residência	19	UF	20	Município de Residência	Código (IBGE)
	21		Distrito		
	22	Bairro	23	Logradouro (rua, avenida,...)	Código
	24	Número	25	Complemento (apto., casa,...)	26
	27		Geo campo 1		
	28		Geo campo 2		
29		Ponto de Referência			
30		CEP			
31		Zona			
		<input type="checkbox"/> 1 - Urbana <input type="checkbox"/> 2 - Rural <input type="checkbox"/> 3 - Periurbana <input type="checkbox"/> 9 - Ignorado			
32		País (se residente fora do Brasil)			
Dados Complementares					
Dados da Pessoa Atendida	33		Nome Social		
	34		Ocupação		
	35		Situação conjugal / Estado civil		
			<input type="checkbox"/> 1 - Solteiro <input type="checkbox"/> 2 - Casado/união consensual <input type="checkbox"/> 3 - Viúvo <input type="checkbox"/> 4 - Separado <input type="checkbox"/> 8 - Não se aplica <input type="checkbox"/> 9 - Ignorado		
	36		Orientação Sexual		
			<input type="checkbox"/> 1- Heterossexual <input type="checkbox"/> 3- Bissexual <input type="checkbox"/> 8- Não se aplica <input type="checkbox"/> 2- Homossexual (gay/lésbica) <input type="checkbox"/> 9- Ignorado		
37		Identidade de gênero:			
		<input type="checkbox"/> 1- Travesti <input type="checkbox"/> 3- Homem Transsexual <input type="checkbox"/> 2- Mulher Transsexual <input type="checkbox"/> 8- Não se aplica <input type="checkbox"/> 9- Ignorado			
38		Possui algum tipo de deficiência/transorno?			
		<input type="checkbox"/> 1- Sim <input type="checkbox"/> 2- Não <input type="checkbox"/> 8- Não se aplica <input type="checkbox"/> 9- Ignorado			
39		Se sim, qual tipo de deficiência/transorno?			
		<input type="checkbox"/> Deficiência Física <input type="checkbox"/> Deficiência visual <input type="checkbox"/> Transtorno mental <input type="checkbox"/> Deficiência Intelectual <input type="checkbox"/> Deficiência auditiva <input type="checkbox"/> Transtorno de comportamento			
Dados da Ocorrência	40	UF	41	Município de ocorrência	Código (IBGE)
	42		Distrito		
	43	Bairro	44	Logradouro (rua, avenida,...)	Código
	45	Número	46	Complemento (apto., casa,...)	47
	48		Geo campo 3		
	49		Geo campo 4		
50		Ponto de Referência			
51		Zona			
		<input type="checkbox"/> 1 - Urbana <input type="checkbox"/> 2 - Rural <input type="checkbox"/> 3 - Periurbana <input type="checkbox"/> 9 - Ignorado			
52		Local de ocorrência			
		<input type="checkbox"/> 01 - Residência <input type="checkbox"/> 04 - Local de prática esportiva <input type="checkbox"/> 07 - Comércio/serviços <input type="checkbox"/> 02 - Habitação coletiva <input type="checkbox"/> 05 - Bar ou similar <input type="checkbox"/> 08 - Indústrias/construção <input type="checkbox"/> 03 - Escola <input type="checkbox"/> 06 - Via pública <input type="checkbox"/> 09 - Outro <input type="checkbox"/> 99 - Ignorado			
53		Ocorreu outras vezes?			
		<input type="checkbox"/> 1 - Sim <input type="checkbox"/> 2 - Não <input type="checkbox"/> 9 - Ignorado			
54		A lesão foi autoprovocada?			
		<input type="checkbox"/> 1 - Sim <input type="checkbox"/> 2 - Não <input type="checkbox"/> 9 - Ignorado			

SVS 15.06.2015

Figura 7 – fichapag1

Violência	55 Essa violência foi motivada por: 01-Sexismo 02-Homofobia/Lesbofobia/Bifobia/Transfobia 03-Racismo 04-Intolerância religiosa 05-Xenofobia 06-Conflito geracional 07-Situação de rua 08-Deficiência 09-Outros 88-Não se aplica 99-Ignorado		
	56 Tipo de violência 1-Sim 2-Não 9-Ignorado <input type="checkbox"/> Física <input type="checkbox"/> Tráfico de seres humanos <input type="checkbox"/> Intervenção legal <input type="checkbox"/> Psicológica/Moral <input type="checkbox"/> Financeira/Econômica <input type="checkbox"/> Outros <input type="checkbox"/> Tortura <input type="checkbox"/> Negligência/Abandono <input type="checkbox"/> Trabalho infantil <input type="checkbox"/> Sexual		
Violência Sexual	57 Meio de agressão 1-Sim 2-Não 9-Ignorado <input type="checkbox"/> Força corporal/espáncamento <input type="checkbox"/> Obj. perfuro-cortante <input type="checkbox"/> Arma de fogo <input type="checkbox"/> Enforcamento <input type="checkbox"/> Substância/Obj. quente <input type="checkbox"/> Ameaça <input type="checkbox"/> Obj. contundente <input type="checkbox"/> Envenenamento, Intoxicação		
	58 Se ocorreu violência sexual, qual o tipo? 1-Sim 2-Não 8-Não se aplica 9-Ignorado <input type="checkbox"/> Assédio sexual <input type="checkbox"/> Estupro <input type="checkbox"/> Pornografia infantil <input type="checkbox"/> Exploração sexual <input type="checkbox"/> Outros		
Dados do provável autor da violência	59 Procedimento realizado 1-Sim 2-Não 8-Não se aplica 9-Ignorado <input type="checkbox"/> Profilaxia DST <input type="checkbox"/> Profilaxia Hepatite B <input type="checkbox"/> Coleta de sêmen <input type="checkbox"/> Contracepção de emergência <input type="checkbox"/> Profilaxia HIV <input type="checkbox"/> Coleta de sangue <input type="checkbox"/> Coleta de secreção vaginal <input type="checkbox"/> Aborto previsto em lei		
	60 Número de envolvidos 1-Sim 2-Não 9-Ignorado 1 - Um <input type="checkbox"/> 2 - Dois ou mais <input type="checkbox"/> 9 - Ignorado <input type="checkbox"/>		
Encaminhamento	61 Vínculo/grau de parentesco com a pessoa atendida 1-Sim 2-Não 9-Ignorado <input type="checkbox"/> Pai <input type="checkbox"/> Ex-Cônjuge <input type="checkbox"/> Amigos/conhecidos <input type="checkbox"/> Policial/agente da lei <input type="checkbox"/> Mãe <input type="checkbox"/> Namorado(a) <input type="checkbox"/> Desconhecido(a) <input type="checkbox"/> Própria pessoa <input type="checkbox"/> Padrasto <input type="checkbox"/> Ex-Namorado(a) <input type="checkbox"/> Cuidador(a) <input type="checkbox"/> Outros <input type="checkbox"/> Madrasa <input type="checkbox"/> Filho(a) <input type="checkbox"/> Patrão/chefe <input type="checkbox"/> Pessoa com relação institucional <input type="checkbox"/> Cônjuge <input type="checkbox"/> Irmão(ã)		
	62 Sexo do provável autor da violência 1-Sim 2-Não 9-Ignorado 1 - Masculino <input type="checkbox"/> 2 - Feminino <input type="checkbox"/> 3 - Ambos os sexos <input type="checkbox"/> 9 - Ignorado <input type="checkbox"/>		
Dados finais	63 Suspeita de uso de álcool 1-Sim 2-Não 9-Ignorado 1-Sim <input type="checkbox"/> 2-Não <input type="checkbox"/> 9-Ignorado <input type="checkbox"/>		
	64 Ciclo de vida do provável autor da violência: <input type="checkbox"/> 1-Criança (0 a 9 anos) 3-Jovem (20 a 24 anos) 5-Pessoa idosa (60 anos ou mais) 2-Adolescente (10 a 19 anos) 4-Pessoa adulta (25 a 59 anos) 9-Ignorado		
Informações complementares e observações	65 Encaminhamento: 1-Sim 2-Não 9-Ignorado <input type="checkbox"/> Rede da Saúde (Unidade Básica de Saúde, hospital, outras) <input type="checkbox"/> Conselho do Idoso <input type="checkbox"/> Delegacia de Atendimento à Mulher <input type="checkbox"/> Rede da Assistência Social (CRAS, CREAS, outras) <input type="checkbox"/> Delegacia de Atendimento ao Idoso <input type="checkbox"/> Outras delegacias <input type="checkbox"/> Rede da Educação (Creche, escola, outras) <input type="checkbox"/> Centro de Referência dos Direitos Humanos <input type="checkbox"/> Justiça da Infância e da Juventude <input type="checkbox"/> Rede de Atendimento à Mulher (Centro Especializado de Atendimento à Mulher, Casa da Mulher Brasileira, outras) <input type="checkbox"/> Ministério Público <input type="checkbox"/> Defensoria Pública <input type="checkbox"/> Conselho Tutelar <input type="checkbox"/> Delegacia Especializada de Proteção à Criança e Adolescente		
	66 Violência Relacionada ao Trabalho 1-Sim 2-Não 9-Ignorado <input type="checkbox"/> 67 Se sim, foi emitida a Comunicação de Acidente do Trabalho (CAT) 1-Sim 2-Não 8-Não se aplica 9-Ignorado <input type="checkbox"/> 68 Circunstância da lesão CID 10 - Cap XX <input type="checkbox"/>		
Notificador	69 Data de encerramento <input type="checkbox"/>		
	Nome do acompanhante _____ Vínculo/grau de parentesco _____ (DDD) Telefone _____		
Observações Adicionais: _____ _____ _____			
Disque Saúde - Ouvidoria Geral do SUS 136 TELEFONES ÚTEIS Central de Atendimento à Mulher 180 Disque Direitos Humanos 100			
Município/Unidade de Saúde _____ Cód. da Unid. de Saúde/CNES _____ Nome _____ Função _____ Assinatura _____			
Violência interpessoal/autoprovocada Sinan SVS 15.06.2015			

Figura 8 – fichapg2

B código

Código R

```

1 library(read.dbc)
2
3 Viol20 <- read.dbc("VIOLBR20.dbc")
4
5 library(dplyr)
6
7 viol2020 <- Viol20 %>%
8   select(DT_NOTIFIC, SG_UF, NU_IDADE_N, CS_SEXO, CS_RACA, SIT_CONJUG, ORIENT
9     _SEX, IDENT_GEN, OUT_VEZES,
10     VIOL_MOTIV, VIOL_FISIC, VIOL_PSICO, VIOL_SEXU, VIOL_FINAN, VIOL_TORT,
11     VIOL_TRAF,
12     VIOL_NEGLI, VIOL_INFAN, VIOL_LEGAL, VIOL_OUTR, VIOL_ESPEC, REL_CONHEC
13     , REL_PAI, REL_MAE,
14     REL_PAD, REL_MAD, REL_CONJ, REL_EXCON, REL_NAMO, REL_EXNAM, REL_FILHO
15     , REL_IRMAO, REL_CUIDA,
16     REL_PATRAO, REL_INST, REL_POL, REL_PROPRI, REL_OUTROS, REL_DESCO, NUM
17     _ENVOLV)
18
19 rm(Viol20)
20
21 save.image("TCC_01AGO-dadosfiltrados.RData")
22
23 violmulheres_20 <- subset(viol2020, CS_SEXO == "F")
24
25 viol_RS_20 <- subset(violmulheres_20, SG_UF == 43)
26 summary(viol_RS_20)
27 summary(viol_RS_20$VIOL_ESPEC)
28
29 viol_RS_20$VIOL_MOTIV = as.factor(viol_RS_20$VIOL_MOTIV)
30 viol_RS_20 <- viol_RS_20 %>%
31   mutate(
32     # Extraíndo o primeiro dgito para identificar o tipo de idade
33     tipo_idade = substr(NU_IDADE_N, 1, 1),
34     # Extraíndo os ltimos dois dgitos como idade real (para quem est em anos
35     )
36     idade_anos = as.numeric(substr(NU_IDADE_N, 2, 4))
37   ) %>%
38   # Filtrar apenas os casos onde o tipo de idade '4' (representando anos)
39   filter(tipo_idade == "4" & idade_anos >= 18)
40

```

```

35 is.numeric(viol_RS_20$idade_anos)
36 summary(viol_RS_20)
37
38 # Transformar as variáveis tipo de violência em uma só coluna
39 viol_RS_20 <- viol_RS_20 %>%
40   mutate(
41     tipo_violencia = paste(
42       ifelse(VIOL_FISIC %in% 1, "Física", NA),
43       ifelse(VIOL_PSICO %in% 1, "Psicológica", NA),
44       ifelse(VIOL_SEXU %in% 1, "Sexual", NA),
45       ifelse(VIOL_FINAN %in% 1, "Financeira", NA),
46       ifelse(VIOL_TORT %in% 1, "Tortura", NA),
47       ifelse(VIOL_NEGLI %in% 1, "Negligência", NA),
48       ifelse(VIOL_OUTR %in% 1, "Outras", NA),
49       sep = ", "
50     ),
51     # Remover vírgulas extras
52
53   )
54 )
55 is.numeric(viol_RS_20$num_envolvidos)
56 summary(as.factor(viol_RS_20$motivo_violencia))
57 #####
58 # Removendo as colunas originais que não são mais necessárias
59 viol_RS_20 <- subset(viol_RS_20, select = -c(CS_RACA, VIOL_MOTIV, IDENT_GEN
60     , ORIENT_SEX, SIT_CONJUG,
61     REL_PAI, REL_MAE, REL_PAD, motiv, REL_MAD,
62     REL_CONJ,
63     REL_EXCON, REL_NAMO, REL_EXNAM, REL_FILHO,
64     REL_IRMAO,
65     REL_CONHEC, REL_DESCO, REL_CUIDA, REL_
66     PATRAO, REL_INST,
67     REL_POL, REL_PROPRI, REL_OUTROS, VIOL_
68     FISIC, VIOL_PSICO,
69     VIOL_SEXU, VIOL_FINAN, VIOL_TORT, VIOL_
70     NEGLI, VIOL_OUTR,
71     DT_NOTIFIC, SG_UF, NU_IDADE_N, CS_SEXO,
72     OUT_VEZES,
73     VIOL_TRAFA, VIOL_INFAN, VIOL_LEGAL, NUM_
74     ENVOLV, tipo_idade,
75     VIOL_ESPEC))
76 #####
77 summary(viol_RS_20)
78
79 table(as.factor(viol_RS_20$VIOL_ESPEC))
80
81 saveRDS(viol_RS_20, "viol_RS_20")

```

```
74
75 write.csv(viol_RS_20, file = "/Users/LEA/Desktop/Alana/viol_RS_20.csv")
76
77 install.packages("openxlsx")
78 library(openxlsx)
79 viol_RS_20 <- na.omit(viol_RS_20)
80 write.xlsx(viol_RS_20, file = "v_RS.xlsx")
81
82 # Remover linhas indesejadas e limpar variáveis específicas
83 viol_RS_20 <- viol_RS_20 %>%
84   filter(!is.na(tipo_violencia)) %>%
85   filter(!is.na(relacao_agressor) & relacao_agressor != "NA") %>%
86   filter(raca != "No especificado") %>%
87   filter(!is.na(relacao_agressor))
88 # Verificar o resumo para confirmar que as filtrações foram aplicadas
89 summary(viol_RS_20)
90 viol_RS_20 <- na.omit(viol_RS_20)
91 summary(viol_RS_20)
92 summary(viol_RS_20$relacao_agressor)
93 summary(viol_RS_20$tipo_violencia)
94 # Salvar o arquivo em Excel após limpeza
95 write.xlsx(viol_RS_20, file = "vi06611_RS.xlsx")
96
97 write.csv2(vi06611_RS, "vio6611_RS.csv")
98
99 dados finais_30Jan2025.csv
100
101 violencia <- read.csv2("vio6611_RS.csv", header = TRUE, sep = ";", dec = ",
102   ",
103   encoding = "UTF-8", stringsAsFactors = FALSE, row.names =
104     1)
105 is.matrix(violencia)
106
107 violencia$idade_anos <- factor(violencia$idade_anos)
108 violencia$raca <- factor(violencia$raca)
109 violencia$tipo_violencia <- factor(violencia$tipo_violencia)
110 violencia$relacao_agressor <- factor(violencia$relacao_agressor)
111 violencia$orientacao_sexual <- factor(violencia$orientacao_sexual)
112 violencia$num_envolvidos <- factor(violencia$num_envolvidos)
113 violencia$situacao_conjugal <- factor(violencia$situacao_conjugal)
114 is.matrix(violencia)
115 is.factor(violencia$idade_anos)
116
117 violencia2 <- data.matrix(violencia)
118 is.matrix(violencia2)
119 summary(violencia2)
```

```
119 library(cluster)
120 indgower <- daisy(dados_vio, metric = "gower", stand = FALSE)
121 indgower
122 # Ward
123 x <- hclust(indgower, method = "ward", members = NULL)
124 plot(x, hang = -1)
125 # Nmero de grupos
126 library(factoextra)
127 # Elbow method
128 fviz_nbclust(violencia, FUN = hcut, method = "wss") +
129   labs(subtitle = "Elbow method")
130 # Silhouette method
131 fviz_nbclust(violencia, FUN = hcut, method = "silhouette") +
132   labs(subtitle = "Silhouette method")
133
134 library(NbClust)
135 grukl <- NbClust(data = indgower, distance = "NULL", min.nc = 2, max.nc =
136   10,
137   method = "ward.D2", index = "kl")
138 grukl
139 # Testar para 20
140 library(dendroextras)
141 gr1 <- slice(x, k = 3)
142 gr1
143 str(gr1)
144 write.csv2(gr1, "agrupamento2.csv")
145
146 res.hcpc <- HCPC(res.famd, nb.clust = -1)
147 temp <- dados_vio %>%
148   mutate_at(c('tipo_violencia', 'relacao_agressor', 'raca', 'orientacao_sexual',
149     'situacao_conjugal'), as.factor)
150 str(temp)
151 dados_vio <- na.omit(temp[, -1])
152
153 # Clusterizao hierrquica
154 cluster_hierarquico <- hclust(matriz_dissimilaridade, method = "ward.D2")
155 plot(cluster_hierarquico, cex = 0.6, hang = -1)
156 # Cortar dendrograma
157 grupos <- slice(cluster_hierarquico, k = 3)
158 # Adicionar clusters aos dados
159 dados_vio$cluster <- grupos
160 summary(grupos)
161
162 ## Plot
163 fviz_cluster(res.hcpc, geom = "point")
```

```
164 library(dplyr)
165 g1 <- filter(res.hcpc$data.clust, clust == 1)
166 g2 <- filter(res.hcpc$data.clust, clust == 2)
167 g3 <- filter(res.hcpc$data.clust, clust == 3)
168 library(autoEDA)
169
170 ## autoEDA
171 auto_res <- autoEDA(res.hcpc$data.clust, y = "clust",
172                    returnPlotList = TRUE,
173                    plotCategorical = "groupedBar",
174                    bins = 30, rotateLabels = TRUE,
175                    color = "#26A69A", verbose = FALSE)
176
177 ## Plot
178 auto_res$plots
179
180 library(xtable)
181 xtable(summary(g1))
182 summary(g1)
183 xtable(summary(g2))
184 summary(g2)
185 xtable(summary(g3))
186 summary(g3)
```



Universidade Federal do Rio Grande – FURG

Instituto de Matemática, Estatística e Física

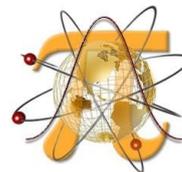
Curso de Bacharelado em Matemática Aplicada

Av. Itália km 8 Bairro Carreiros

Rio Grande-RS CEP: 96.203-900 Fone (53)3293.5411

e-mail: imef@furg.br

Sítio: www.imef.furg.br



Ata de Defesa de Monografia

No terceiro dia do mês de fevereiro de 2025, às 09h30min, no Auditório do IMEF, foi realizada a defesa do Trabalho de Conclusão de Curso da acadêmica **Alana Baldez de Avila**, intitulada “**Violência de Gênero Durante a Pandemia: Identificação de Padrões em Notificações do Rio Grande do Sul Usando Técnicas de Agrupamento de Dados Mistos**”, sob a orientação do Prof. Dr. Juliano C. Marangoni e co-orientação da Profa. Dra. Raquel da Fontoura Nicolette, ambos deste instituto. A banca avaliadora foi composta pela Profa. Dra. Débora Spenassato e pelo Prof. Dr. Felipe Ricardo Santos de Gusmão, ambos do IMEF/FURG. A candidata foi: (X) aprovada por unanimidade; () aprovada somente após satisfazer as exigências que constam na folha de modificações, no prazo fixado pela banca; () reprovada. Na forma regulamentar, foi lavrada a presente ata, que é abaixo assinada pelos membros da banca, na ordem acima relacionada.

Documento assinado digitalmente



JULIANO CESAR MARANGONI

Data: 03/02/2025 15:17:27-0300

Verifique em <https://validar.iti.gov.br>

Prof. Dr. Juliano C. Marangoni
Orientador

Documento assinado digitalmente



RAQUEL DA FONTOURA NICOLETTE

Data: 12/02/2025 13:56:01-0300

Verifique em <https://validar.iti.gov.br>

Profa. Dra. Raquel da Fontoura Nicolette
Co-Orientadora

Documento assinado digitalmente



DEBORA SPENASSATO

Data: 03/02/2025 15:28:54-0300

Verifique em <https://validar.iti.gov.br>

Profa. Dra. Débora Spenassato

Documento assinado digitalmente



FELIPE RICARDO SANTOS DE GUSMAO

Data: 03/02/2025 16:32:04-0300

Verifique em <https://validar.iti.gov.br>

Prof. Dr. Felipe Ricardo Santos de Gusmão