

Explorando Dados Governamentais para Prevenção da Violência de Gênero: Uma Abordagem Visual

Eduarda Patricio¹, Gabriel Zurawski¹, André Rollwagen^{1,2}, Isabel Manssour¹

¹School of Technology, DaVInt – *Data Visualization and Interaction Lab*
Pontifícia Universidade Católica do Rio Grande do Sul – (PUCRS)
Porto Alegre – RS – Brasil

²Instituto Federal de Educação, Ciência e Tecnologia Sul-rio-grandense (IFSUL)
Passo Fundo – RS – Brasil
{e.patricio, zurawski.gabriel, andre.rollwagen}@edu.pucrs.br,
isabel.manssour@pucrs.br

Abstract. *Data visualization facilitates the identification of patterns and trends, providing insights that help with decision-making. Governments have made open datasets available, including cases of violence against women. Gender-based violence is considered a public health and safety problem and requires approaches that help identify behavior patterns, trends, and profiles of victims and aggressors. This paper proposes the application of interactive visualization techniques to explore government data on violence against women, aiming for better understanding and preventive actions. We demonstrate how these techniques reveal hidden patterns and help identify vulnerabilities, with the purpose of developing strategies for preventing and combating gender-based violence.*

Resumo. *A visualização de dados facilita a identificação de padrões e tendências, fornecendo insights que ajudam na tomada de decisões. Governos têm disponibilizado dados abertos, incluindo casos de violência contra a mulher. A violência de gênero é considerada um problema de saúde e segurança pública e requer abordagens que ajudem a identificar padrões de comportamento, tendências e perfis das vítimas e dos agressores. Este artigo propõe a aplicação de técnicas de visualização interativa para explorar dados governamentais sobre violência contra a mulher, visando melhor compreensão e ações preventivas. Mostramos como elas revelam padrões ocultos e ajudam a identificar vulnerabilidades, buscando prevenir e combater a violência de gênero.*

1. Introdução

O aumento na geração de dados, estimulado pelo progresso tecnológico, destaca a importância de estruturar esses dados para a análise e obtenção de informações relevantes [Pentland 2013]. [Liu et al. 2017] afirmam que a visualização desempenha um papel importante no processo de análise de dados, pois permite uma melhor compreensão das informações contidas nos conjuntos de dados. [Tufte 2001] destaca que uma visualização eficaz facilita a identificação de tendências, correlações, padrões e discrepâncias. Assim, facilita a obtenção de *insights* e a tomada de decisão [Kirk 2016].

Atualmente, os governos têm se esforçado para promover transparência ao disponibilizar uma vasta gama de dados abertos que podem ser analisados. Esses dados

são oriundos de diversas áreas e setores, incluindo denúncias relacionadas à violência contra a mulher. Neste caso, os dados fornecem informações sobre os registros realizados, e contemplam o tipo de violência, o local onde ocorreu, o perfil da vítima e do agressor, além de dados demográficos relevantes. A violência contra a mulher é um problema de segurança e de saúde pública que afeta a sociedade como um todo [de Almeida Teles and de Melo 2017]. É uma manifestação de desigualdade de gênero e violação dos direitos humanos, que pode causar danos físicos, psicológicos e sociais. É muito importante abordar essa questão de forma abrangente, compreendendo suas origens e consequências, para promover uma mudança cultural e estrutural que elimine a tolerância à violência de gênero, além de proteger as vítimas.

A análise de dados relacionados à violência contra a mulher tem sido explorada na literatura. Por exemplo, [Júnior and Ribeiro 2019] conduziram um mapeamento da literatura para identificar estudos que fazem uma análise preditiva da violência doméstica. Eles descobriram fatores relevantes para a predição, como idade, histórico de maus-tratos na infância, consumo de álcool e influência da cultura local. [Montenegro 2021] empregou técnicas de coleta, análise e visualização de dados abertos do governo para investigar indicadores de violência contra mulheres, antes e durante a pandemia de COVID-19. [Rios et al. 2023] integraram diversos conjuntos de dados para examinar informações sociodemográficas e criminais sobre homicídio de meninas e mulheres em Porto Alegre.

Embora existam iniciativas para coleta, disponibilização e análise de dados abertos do governo relacionados à violência contra a mulher, ainda há espaço para pesquisa nesta área, visando facilitar a exploração e acessibilidade dos dados para pesquisadores, profissionais e formuladores de políticas públicas. Neste contexto, a análise exploratória visual destes dados abertos pode desempenhar um papel importante, facilitando a obtenção de *insights* por meio da identificação de padrões, tendências e correlações ocultas.

Este trabalho tem como objetivo mostrar como o uso de técnicas de visualização interativa, combinadas com métodos de exploração de dados, ajuda a conhecer e obter *insights* relevantes de dados abertos do governo relacionados à violência contra a mulher. Para isso, foram desenvolvidos *dashboards* que fomentam um melhor conhecimento sobre estes dados e auxiliam na identificação das variáveis mais relevantes, visando apoiar o projeto de ações preventivas. Nossa pesquisa se destaca pela interatividade das visualizações criadas e pela facilidade com que os usuários podem recriá-las. Disponibilizamos no *GitHub*¹ um guia passo a passo de como coletar e processar os dados para posterior visualização, permitindo que outros pesquisadores reproduzam e adaptem nossas análises de forma transparente e acessível.

O restante deste artigo está organizado da seguinte maneira: na Seção 2, exploramos alguns trabalhos relacionados. Em seguida, detalhamos a ferramenta para análise visual desenvolvida. Na Seção 4, apresentamos as discussões, descrevendo nossas contribuições, os *insights* obtidos e as limitações do trabalho. Na última seção descrevemos nossas conclusões e as possibilidades de pesquisas futuras.

2. Trabalhos Relacionados

Um estudo da literatura sobre o uso de aprendizado de máquina para facilitar a análise de dados de violência doméstica, buscou destacar os dados mais relevantes para análise pre-

¹<https://github.com/DAVINTLAB/GenderViolence-DataVis>

ditiva [Júnior and Ribeiro 2019]. Dentre eles, foram evidenciados a idade, o uso de álcool, a cultura local, as questões econômicas e o nível de escolaridade. [Montenegro 2021] empregou técnicas de coleta, análise e representação visual de dados para examinar os indicadores de violência contra mulheres, antes e durante o período de distanciamento social da pandemia de COVID-19, visando compreender os impactos da pandemia nestes casos de violência e gênero no Brasil. Os resultados mostraram que houve uma redução nos casos de violência, mas que pode ter sido influenciada pelo aumento de subnotificações.

[Rios et al. 2023] analisaram dados sociodemográficos, criminais e médico-legais relacionados aos casos de homicídio de meninas e mulheres na cidade de Porto Alegre-RS, de 2010 a 2016. O estudo mostra um aumento nas taxas gerais de homicídios femininos, especialmente devido à violência urbana, como envolvimento com o tráfico de drogas, crimes e roubos. Mulheres não brancas entre 15 e 29 anos foram as mais afetadas, e os autores afirmam que a obtenção de dados detalhados sobre as vítimas e as características da violência é essencial para enfrentar o problema e orientar políticas públicas.

Dados do VIVA Inquérito de 2017 ² foram usados para analisar a violência física cometida por parceiros íntimos contra mulheres que buscaram atendimento nos serviços públicos de urgência e emergência no Brasil [Vasconcelos et al. 2022]. Foram examinadas as características sociodemográficas das vítimas e agressores, e aspectos da natureza da violência. Os resultados revelaram padrões associados a variáveis como raça, idade, nível de escolaridade das vítimas, forma de agressão, entre outros aspectos relevantes.

[Leite et al. 2023] utilizaram dados de notificação de casos de violência contra mulheres no estado do Espírito Santo, entre 2011 e 2018. O objetivo era analisar a recorrência da violência em diferentes estágios da vida feminina e investigar os fatores associados. As análises estatísticas revelaram uma significativa incidência de violência repetitiva dirigida a mulheres, destacando-se os casos envolvendo idosas com deficiência ou transtornos. Os agressores, a maioria com cerca de 25 anos, eram principalmente familiares das vítimas.

Uma análise de fatalidades de mulheres entre 15 e 59 anos em cidades do Paraná, com o objetivo de estimar o risco de violência letal e identificar fatores correlacionados, foi realizada por [Moroskoski et al. 2022]. Eles encontraram uma relação entre violência letal e a proporção de mães que são chefes de família. Além disso, observaram uma associação entre taxas de violência não letal e a presença de prefeitas e vereadoras nas cidades estudadas. As análises mostraram que a violência está associada a baixos níveis educacionais, violência estrutural e participação política das mulheres.

A falta de completude dos dados e o uso de dados secundários que podem ter subnotificação de casos de violência, é uma limitação destacada nos trabalhos [Leite et al. 2023, Moroskoski et al. 2022]. Outra limitação é a apresentação restrita dos dados sobre o perfil das vítimas, devido ao pequeno número de estudos envolvendo variáveis de segurança pública e ao desafio de conciliar dados de diferentes bancos de dados [Rios et al. 2023].

Estes trabalhos, sintetizados na Tabela 2, mostram a diversidade de abordagens usadas na análise de dados relacionados à violência doméstica e de gênero. Mas, também é possível observar uma limitação relacionada à abrangência geográfica dos dados ana-

²<https://www.gov.br/saude/pt-br/composicao/svsa/inqueritos-de-saude/viva-inquerito>

lisados, que referem-se apenas a um município [Montenegro 2021, Rios et al. 2023] ou estado [Moroskoski et al. 2022]. [Vasconcelos et al. 2022] usaram dados de serviços de urgência e emergência do SUS de 23 capitais e alguns municípios do país. Assim, a amostra retrata apenas a população atendida nesses serviços específicos. Além disso, nem todos usam dados públicos, dificultando a reprodução ou extensão da pesquisa, e não disponibilizam uma ferramenta para analisar os dados.

Tabela 1. Estudos envolvendo violência doméstica e de gênero.

Referência	Ano	Objetivos	Dados usados
Pereira Júnior e Ribeiro	2019	Destacar dados relevantes para análise preditiva de violência doméstica.	Dados extraídos da revisão da literatura.
Montenegro	2021	Compreender as consequências da pandemia nos casos de violência de gênero no Brasil.	Dados abertos do governo.
Vasconcelos et al.	2022	Investigar a violência física por parceiros íntimos contra mulheres, registrada em serviços de urgência e emergência no Brasil.	VIVA Inquérito de 2017.
Rios et al.	2023	Analisar informações sociodemográficas e criminais dos homicídios femininos em Porto Alegre.	Dados privados.
Leite et al.	2023	Analisar a recorrência da violência ao longo das diferentes fases da vida feminina no Espírito Santo.	Dados de notificação de casos de violência contra mulheres.
Moroskoski et al.	2023	Estimar o risco associado à violência letal contra mulheres e identificar fatores correlacionados em cidades do Paraná.	Dados sobre fatalidades com mulheres entre 15 e 59 anos.

3. Ferramenta para Análise Visual

Nesta seção é descrita a ferramenta desenvolvida, que usa técnicas de visualização interativas que permitem analisar e obter *insights* de dados abertos do governo relacionados à violência contra a mulher. Inicialmente, foi feita a coleta e pré-processamento dos dados. Depois, então, foram elaborados *dashboards* com as visualizações interativas.

Para a preparação dos dados foram utilizados o *Tableau Prep Builder* e a linguagem *Python*. O primeiro foi usado para processar e organizar em um único arquivo diversas etapas de processamento de dados, como a limpeza e união de linhas ou colunas de conjuntos de dados. A linguagem *Python* possui diversas bibliotecas, tais como *Pandas* e *NumPy*, que permitiram manipular os dados e aplicar métodos estatísticos. Para elaboração dos *dashboards* foi usado o *Tableau Desktop*, que permite gerar visualizações de maneira rápida e interativa, e organizá-las em *dashboards*. Além disso, todos os *dashboards* desenvolvidos podem ser compartilhados no *Tableau Public*.

3.1. Coleta e Preparação dos Dados

Inicialmente, foram coletados os dados do portal de dados abertos do Ministério dos Direitos Humanos e da Cidadania³. O portal disponibiliza 16 arquivos *.csv*, organizados por ano ou semestre desde 2014, mas apenas 15 desses arquivos foram utilizados neste trabalho. O arquivo “Balanço 2019” foi excluído da análise, pois continha apenas um resumo do ano de 2019, já com sua tabela consolidada. Os dados contidos nos arquivos incluem

³<https://www.gov.br/mdh/pt-br/aceso-a-informacao/dados-abertos/ligue180>

informações como cidade, estado, horário da denúncia, identidade do denunciante, se a vítima estava em risco de vida, local da violência, além de dados demográficos como faixa etária, escolaridade, faixa de renda e profissão da vítima e do suspeito, entre outros. O pré-processamento desses dados foi dividido em cinco etapas, descritas a seguir. Alguns dos processamentos precisaram ser feitos através de *scripts* em *Python*, não sendo possível realizar diretamente pelo *Tableau Prep Builder*. Além disso, para análises relativas ao tamanho da população (por estado, município, raça, e grau de instrução), também foram utilizados os dados do Censo de 2022⁴.

- **1º Etapa:** filtragem por violência contra mulher e conversão dos arquivos para *.xlsx*. Os dados a partir de 2020 incluem denúncias não apenas sobre violência doméstica, mas também sobre crimes contra o meio ambiente ou contra idosos, por exemplo. Por isso, é feita uma filtragem por dados exclusivamente sobre violência doméstica. Além disso, por virem em formato *.csv*, os arquivos com os dados passam por uma conversão para *.xlsx* buscando reduzir seu tamanho e acelerar futuros processamentos. Etapa realizada através do *Tableau Prep Builder*.
- **2º Etapa:** remoção de linhas com *hashs* repetidos. Os dados a partir de 2020 armazenam um código *hash* para cada denúncia. Ao baixar os arquivos, foi possível perceber que muitos desses códigos *hash* eram repetidos, e a quantidade de denúncias era muito maior que em anos anteriores, sinalizando que haviam denúncias repetidas. Portanto, foi feito e executado um *script* em *Python* para remoção das denúncias repetidas.
- **3º Etapa:** união e limpeza de tabelas de mesmos formatos. Ao longo dos anos houveram quatro alterações na forma de coleta e registro de dados das denúncias. Por isso, esta etapa busca fazer a união dessas tabelas com os mesmos formatos e, em seguida, a sua limpeza, removendo valores incoerentes ou colunas com dados faltantes (a união de todos os conjuntos de dados resultantes é feita posteriormente, na 4º etapa). Foram criados quatro conjuntos de dados: um de 2014 a novembro de 2018, outro de dezembro de 2018 a dezembro de 2019, um do primeiro semestre de 2020, e outro do segundo semestre de 2020 até dezembro de 2023. A identificação das tabelas com formatos iguais foi feita por meio de um arquivo *Excel* contendo os nomes das colunas de cada tabela, sendo a união e limpeza realizadas no *Tableau Prep Builder*.
- **4º Etapa:** padronização das profissões, países e municípios. Dados mais recentes possuíam formas diferentes de registrar nomes de municípios, alguns continham códigos antes do nome e isso impedia a criação de gráficos com distribuição geográfica. Portanto, foi feito um *script* em *Python* para a padronização dos nomes. Já as colunas que elencavam as profissões das vítimas e dos suspeitos, funcionavam como campos abertos e tornavam inviável sua análise. Para isso foi incluída no *script* uma padronização das profissões em seis grupos diferentes, baseados no dicionário de dados do ENEM⁵, a qual pode ser vista na Tabela 2.
- **5º Etapa:** união de tabelas de diferentes formatos. Buscando criar um conjunto de dados único para análise, que contemplasse o período de 2014 a 2023, foram unificados os quatro conjuntos de dados organizados anteriormente. Contudo, nesse processo, algumas colunas perderam consistência e representatividade, como, por exemplo, as colunas que detalhavam a escolaridade da vítima e suspeito. Em alguns conjuntos de dados, a escolaridade era apresentada de maneira detalhada, especificando se a etapa

⁴<https://tinyurl.com/y32c879j>

⁵<https://tinyurl.com/2s4f5f2k>

Tabela 2. Classificação de profissões segundo o dicionário de dados do ENEM

Grupo	Profissões
Grupo 1	Lavrador, agricultor sem empregados, bóia fria, criador de animais, apicultor, pescador, lenhador, seringueiro, extrativista.
Grupo 2	Diarista, empregado doméstico, cuidador de idosos, babá, cozinheiro (em casas particulares), motorista particular, jardineiro, faxineiro de empresas e prédios, vigilante, porteiro, carteiro, <i>office-boy</i> , vendedor, caixa, atendente de loja, auxiliar administrativo, recepcionista, servente de pedreiro, repositor de mercadoria.
Grupo 3	Padeiro, cozinheiro industrial ou em restaurantes, sapateiro, costureiro, joalheiro, torneiro mecânico, operador de máquinas, soldador, operário de fábrica, trabalhador da mineração, pedreiro, pintor, eletricista, encanador, motorista, caminhoneiro, taxista.
Grupo 4	Professor (de ensino fundamental ou médio, idioma, música, artes etc.), técnico, policial, militar de baixa patente, corretor de imóveis, supervisor, gerente, mestre de obras, pastor, microempresário, pequeno comerciante, pequeno proprietário de terras, trabalhador autônomo ou por conta própria.
Grupo 5	Médico, engenheiro, dentista, psicólogo, economista, advogado, juiz, promotor, defensor, delegado, tenente, capitão, coronel, professor universitário, diretor em empresas públicas e privadas, político, proprietário de empresas com mais de 10 empregados.

educacional havia sido completada ou não (por exemplo, ensino médio completo ou incompleto). Outros conjuntos de dados limitavam-se a mencionar apenas o grau de escolaridade, sem oferecer detalhes sobre a conclusão dos níveis educacionais. Por isso, parte da representatividade foi perdida. Essa etapa foi realizada com o *Tableau Prep Builder*.

3.2. Perfis Demográficos

Foram desenvolvidos dois *dashboards* com o propósito de identificar os perfis demográficos das vítimas e dos suspeitos, apresentados, respectivamente, nas Figuras 1 e 2. Eles possibilitam explorar dados como raça/cor, nível de instrução, faixa etária e profissão. Além disso, no *dashboard* do perfil da vítima, são incluídas análises sobre a variação da faixa de renda e a distribuição por estado do número de vítimas. Ambos podem ser filtrados com base na profissão da vítima/suspeito, e o *dashboard* do perfil da vítima oferece a funcionalidade de filtragem por estado por meio de um mapa interativo.

3.3. Relações entre Vítimas e Suspeitos

Um terceiro *dashboard*, apresentado na Figura 3, disponibiliza gráficos que permitem visualizar quando iniciaram, a frequência e o cenário das violações, além de esclarecer a relação entre a vítima e o suspeito.

3.4. Distribuição das Denúncias e Vítimas

Para examinar a distribuição geográfica das denúncias e das vítimas, foi criado um quarto *dashboard* que oferece um mapeamento detalhado por estados e municípios (Figura 4). Ele disponibiliza uma visualização precisa da localização das ocorrências e das áreas com maiores incidências de casos, proporcionando *insights* sobre as regiões que apresentam um maior risco para possíveis vítimas. Para calcular a incidência de denúncias e de vítimas em cada estado ou município, dividimos o número de vítimas pela população total da respectiva área. Em seguida, multiplicamos o resultado por cem mil. Essa prática permite uma comparação padronizada entre diferentes regiões, levando em consideração a variação de suas populações. Este *dashboard* também disponibiliza um filtro por ano.

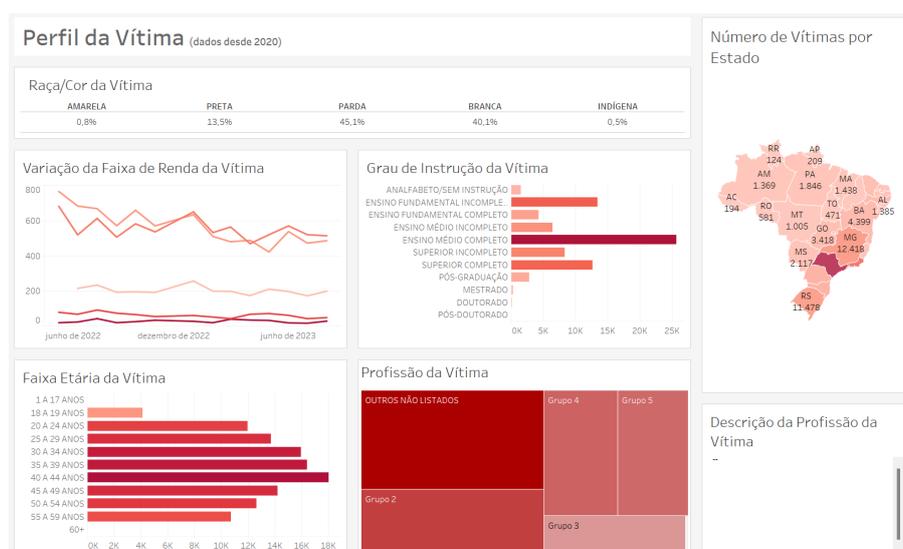


Figura 1. *Dashboard* para analisar o perfil demográfico da vítima.

3.5. Frequência de Denúncias

Com o objetivo de analisar o padrão temporal das denúncias, também foi elaborado um *dashboard* que apresenta a variação do total de denúncias ao longo dos anos. Ele permite comparar a frequência das denúncias entre os meses dos anos e os dias da semana, além de incluir um *heatmap* que correlaciona essas duas variáveis. Assim, é possível determinar padrões nas ocorrências ao longo do tempo. Como mostra a Figura 5, as visualizações permitem ver a dinâmica temporal das denúncias e identificar tendências.

4. Discussão

Todos os *dashboards* desenvolvidos neste trabalho e descritos na Seção 3, estão disponíveis para acesso no *Tableau Public*⁶, sendo possível interagir diretamente com eles e analisar os dados a partir das visualizações fornecidas. Nesta seção, é feita uma discussão sobre os *insights* obtidos através desse estudo e das interações com os *dashboards*.

4.1. Insights Obtidos

Analisando o *dashboard* “*Perfil da Vítima*”, observa-se uma diversidade étnica com predominância de indivíduos identificados como pardos (45,1%) e brancos (40,1%). Nota-se, que embora a raça amarela represente uma menor quantidade de vítimas em números absolutos (0,8%), ela apresenta a maior proporção em relação ao tamanho de sua população (850.130 pessoas), com 0.34%. Quanto à distribuição de renda entre as vítimas, apesar de variar ela não apresenta uma tendência clara ao longo do tempo, sugerindo que o impacto é transversal às classes econômicas. No entanto, é importante observar que a maioria das vítimas pertence a faixas de renda que variam entre 1 e 5 salários mínimos, ou menos. Já a categorização das profissões da vítima, excluindo “Outros não listados”, é majoritariamente listada como Grupo 2, que contempla profissões como, por exemplo, diaristas e empregadas domésticas. Em relação a educação, há uma maior concentração de vítimas

⁶<https://public.tableau.com/app/profile/gabriel.zurawski/viz/GenderViolence-DataVis/Inicio?publish=yes>



Figura 2. Dashboard para analisar o perfil demográfico do suspeito.

com ensino médio completo e ensino fundamental incompleto. Porém, em números relativos à população, a maior parte das vítimas possui ensino médio ou superior incompleto, são pardas e possuem renda de 1 à 5 salários mínimos. Estes dados mostram um perfil muito similar ao encontrado por [Montenegro 2021], reforçando ainda mais a fragilidade de atuação de políticas públicas em áreas de vulnerabilidade social.

No *dashboard* “Perfil do Suspeito”, observa-se uma disparidade significativa de gênero, com o sexo masculino representando a maioria. A distribuição étnica é semelhante à das vítimas, com brancos (42,1%) e pardos (40,7%) sendo os mais recorrentes. Em relação à faixa etária, o grupo mais prevalente é o de 35 a 39 anos. A maioria dos suspeitos possui ensino fundamental incompleto, seguido por ensino médio completo. Devido a quantia muito grande de valores nulos sobre a faixa de renda do suspeito, não foi possível analisá-la. No entanto, a categorização profissional dos suspeitos, excluindo “Outros não listados”, é majoritariamente listada como Grupo 3, que inclui profissões como caminhoneiros e operadores de máquina. Com base nessas informações, podemos deduzir que o suspeito possui uma renda similar ou ligeiramente superior à da vítima.

Ao examinar o período de início das violações no *dashboard* “Relações entre Vítimas e Suspeitos”, destaca-se a tendência de concentração dos casos na categoria “Há mais de um ano”. Isto sugere que as vítimas possam estar mais propensas a denunciar os abusos após um período maior de relacionamento, dado o menor número de denúncias classificadas como “Há mais de cinco anos” e “Há mais de dez anos”. Outro aspecto relevante é a variação na frequência de violações, observando-se um aumento da categoria “Ocasionalmente” ao passar do grupo de companheiros para o grupo de ex-companheiros.

Referente à distribuição de vítimas pelo país, com base em dados coletados a partir de 2020, São Paulo e Rio de Janeiro se destacam pelo maior número absoluto de casos registrados. Contudo, quando se considera a taxa de vítimas relativas a população, as maiores taxas de vítimas são do Rio de Janeiro (45%) e do Distrito Federal (34%). Em termos de denúncias, tanto o número quanto a taxa de denúncias seguem esse mesmo padrão de distribuição.

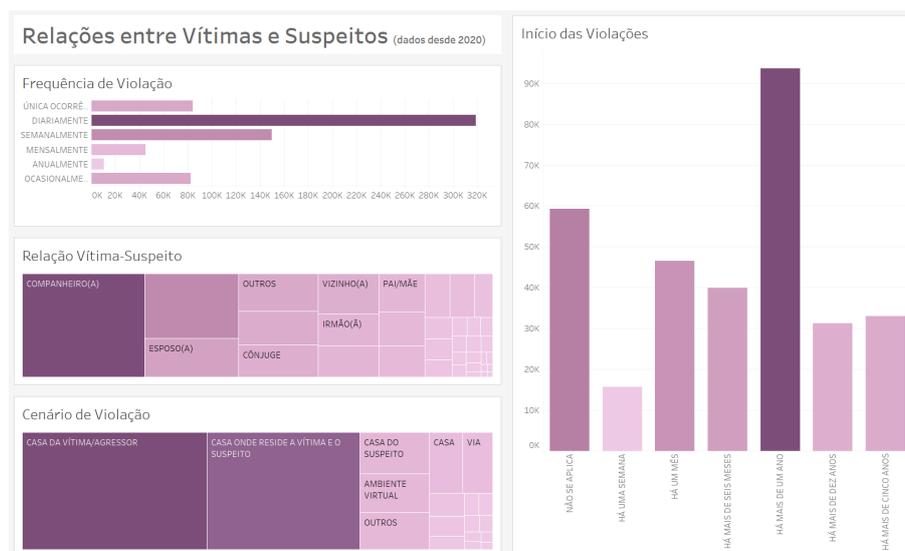


Figura 3. Dashboard para analisar as relações entre Víctimas e Suspeitos.

Ao analisar as distribuições temporais das denúncias apresentadas no *dashboard* “*Frequência de Denúncias*”, é possível identificar um padrão oscilante ao longo dos anos, com picos notáveis entre o quarto trimestre de 2015 e o segundo trimestre de 2018, além de um aumento significativo no segundo trimestre de 2020. Observa-se também uma tendência de crescimento gradual no volume de denúncias em 2022 e 2023, possivelmente indicando uma retomada após um período de declínio. Esse declínio pode estar relacionado aos impactos da pandemia de COVID-19, que inicialmente resultou em um pico seguido por uma redução substancial.

[Montenegro 2021] demonstra que durante a pandemia, em momentos de maior índice de isolamento social, os casos de violência doméstica apresentaram uma pequena queda. Ele destaca um pico de isolamento social em março de 2020, que corresponde também ao aumento de denúncias no segundo trimestre do mesmo ano, apontado pelo nosso gráfico. Após esse período, há uma redução gradual nas denúncias, a qual é também acompanhada pela queda no índice de isolamento social, apresentado por Montenegro.

Em relação à frequência semanal, nota-se que as segundas-feiras registram o maior número de denúncias, seguidas pelas quartas e terças-feiras, sugerindo uma tendência de acúmulo de casos durante o fim de semana que são reportados no início da semana de trabalho. O *heatmap* correlacionando o mês e o dia da semana revela algumas combinações de meses e dias apresentando uma densidade maior de denúncias.

Examinando as denúncias por mês, percebe-se uma distribuição relativamente uniforme ao longo do ano, com variações que podem estar relacionadas a fatores sazonais ou eventos específicos, como feriados ou temporada de férias. Por fim, a análise horária revela um aumento significativo de denúncias nas primeiras horas da manhã, o que pode refletir uma tendência de reportar incidentes logo no início do dia útil.

4.2. Potenciais Usuários

A sociedade como um todo se beneficiaria da disponibilidade de dados com interpretações visuais, tornando a informação mais acessível. Atualmente, o acesso aos dados governa-

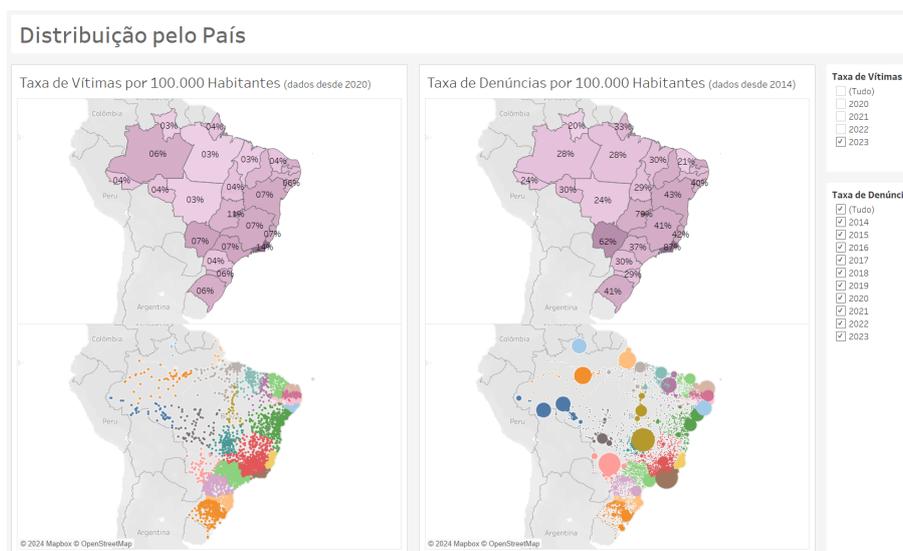


Figura 4. Dashboard da distribuição de denúncias e vítimas pelo país.

mentais é pouco intuitivo, dificultando seu uso pela população em geral. Funcionários que registram dados sobre violência de gênero poderiam usar a ferramenta para identificar padrões rapidamente, melhorando o cadastro e o atendimento oferecido pelo Disque 100 e pelo Ligue 180. Isso promoveria um atendimento mais humanizado e eficiente às vítimas, beneficiando setores como segurança e saúde pública. Membros da segurança pública com acesso a *dashboards* detalhados poderiam identificar áreas com maior incidência de violência, permitindo uma resposta rápida e eficiente, além de auxiliar na elaboração de estratégias de prevenção. Para esta aplicação específica, seria interessante conversar com profissionais da área para avaliar a viabilidade e a utilidade da ferramenta.

4.3. Limitações

A análise dos dados revelou uma grande quantidade de valores nulos, o que tornou desafiadora a interpretação de alguns resultados. Por exemplo, cerca de 90% dos registros da faixa de renda do suspeito eram nulos, impossibilitando sua análise e comparação com a faixa de renda da vítima. Outra dificuldade encontrada durante o estudo foi a variação frequente no formato dos dados fornecidos pelo governo. Essas alterações recorrentes prejudicaram a criação de um conjunto de dados único e padronizado, essencial para uma análise consistente e abrangente.

Nos trabalhos de [Soares and de Medeiros Filho 2020] e [Mendoza et al. 2023] foram feitas revisões e análises da literatura quanto a Governo Inteligente e Governança de Dados, respectivamente. Ambos apresentam a importância da gestão governamental baseada em dados e também problemas relacionados encontrados em outros trabalhos. Eles também apontam a falta de padronização dos dados como um dos principais problemas encontrados ainda hoje, e [Mendoza et al. 2023]. destacam os principais problemas encontrados em dados de portais governamentais, tais como: dados incompletos e ausentes, e preenchimento de forma incorreta. Todos estes problemas ocorreram nos dados utilizados nesse trabalho. A inconsistência nos formatos dos dados ressalta a importância da adoção de padrões e protocolos claros para a divulgação de dados governamentais, visando facilitar análises futuras e garantir que as informações sejam confiáveis.



Figura 5. *Dashboard* para análises temporais das denúncias.

5. Conclusões

Neste estudo, buscamos apresentar um diagnóstico sobre a violência contra a mulher no Brasil, analisando os dados do Disque 100 e do Ligue 180. Também desenvolvemos uma metodologia de código aberto para o processamento e visualização dos dados que pode ser aplicada no futuro, conforme mais dados forem sendo divulgados nos próximos anos. Além disso, foi possível constatar a necessidade de um maior cuidado na coleta e organização dos dados por parte do governo, para viabilizar uma melhor análise de longo prazo através da padronização desses dados. Essa abordagem pode fornecer uma base sólida para futuras investigações e tomadas de decisão. Ao apresentarmos as nossas descobertas de forma clara e objetiva, esperamos ter contribuído para uma melhor compreensão do tema e para o progresso do conhecimento nesta área.

O acompanhamento dos dados divulgados semestralmente pode ser feito de forma mais rápida com a solução proposta, permitindo análises ágeis para diagnósticos mais frequentes sobre a quantidade e os tipos de casos. Como trabalhos futuros, pretendemos avaliar os *dashboards* junto a membros da segurança pública e aprofundar o estudo, investigando as razões pelas quais as denúncias variam ao longo do tempo, o que pode contribuir para a diminuição de subnotificações e para o aperfeiçoamento das técnicas de prevenção e intervenção contra a violência de gênero. Além disso, existem campos nos dados ainda não explorados, como o uso de álcool e drogas, dependência financeira, local de residência, motivação, entre outros. Por fim, é viável aplicar técnicas estatísticas mais avançadas para identificar padrões e correlações com maior precisão, incluindo o uso de aprendizado de máquina para estimar tendências.

Agradecimentos

Patrício, Souza e Manssour agradecem o apoio financeiro, respectivamente, do Programa de Educação Tutorial (PET), da bolsa CNPq PIBITI e da bolsa CNPq de produtividade 303208/2023-6. Também apoiado pelo Ministério da Ciência, Tecnologia e Inovações, com recursos da Lei 8.248, de 23/10/1991, no âmbito do PPI-SOFTEX, coordenado pela Softex e publicado Residência em TIC 02-Aditivo, DOU 01245.012095/2020-56.

Referências

- de Almeida Teles, M. A. and de Melo, M. (2017). *O que é violência contra a mulher*. Primeiros Passos. Brasiliense.
- Júnior, A. P. and Ribeiro, F. (2019). Análises de dados de violência doméstica com o uso de aprendizagem de máquina: um mapeamento sistemático. In *Anais da VII Escola Regional de Computação do Ceará, Maranhão e Piauí*, pages 143–150, Porto Alegre, RS, Brasil. SBC.
- Kirk, A. (2016). *Data Visualization: A Handbook for Data-Driven Design*. SAGE Publications, 1st edition.
- Leite, F., Thomazini Piske Garcia, M., Cavalcante, G., Venturin, B., Pedroso, M., Gomes, E., and Tavares, F. L. (2023). Violência recorrente contra mulheres: análise dos casos notificados. *Acta Paulista de Enfermagem*, 36.
- Liu, S., Maljovec, D., Wang, B., Bremer, P.-T., and Pascucci, V. (2017). Visualizing high-dimensional data: Advances in the past decade. *IEEE Transactions on Visualization and Computer Graphics*, 23(3):1249–1268.
- Mendoza, I., Corrêa, R., and Bernardini, F. (2023). Como a governança de dados pode auxiliar na mitigação de barreiras de uso de portais de dados governamentais abertos? uma análise da literatura. In *Anais do XI Workshop de Computação Aplicada em Governo Eletrônico*, pages 212–223, Porto Alegre, RS, Brasil. SBC.
- Montenegro, T. C. (2021). Violência contra a mulher: Análise comparativa de dados públicos antes e durante a pandemia de covid-19. Monografia (Trabalho de Conclusão de Curso).
- Moroskoski, M., Neto, F. C., Machado de Brito, F. A., Ferracioli, G. V., de Oliveira, N. N., de Carvalho Dutra, A., Baldissera, V. D. A., and de Oliveira, R. R. (2022). Lethal violence against women in southern brazil: Spatial analysis and associated factors. *Spatial and Spatio-temporal Epidemiology*, 43:100542.
- Pentland, A. (2013). The data-driven society. *Scientific American*, 309:78–83.
- Rios, A. M. F. M., Crespo, K. C., Martini, M., Telles, L. E. D. B., and Magalhães, P. V. S. (2023). Gender-related and non-gender-related female homicide in porto alegre, brazil, from 2010 to 2016. *PLOS ONE*, 18(3):1–12.
- Soares, G. and de Medeiros Filho, F. (2020). Governança inteligente com análise e integração de dados: Uma revisão sistemática de literatura. In *Anais do VIII Workshop de Computação Aplicada em Governo Eletrônico*, pages 132–139, Porto Alegre, RS, Brasil. SBC.
- Tufte, E. R. (2001). *The Visual Display of Quantitative Information*. Graphics Press, 2nd edition.
- Vasconcelos, N. M. d., Andrade, F. M. D. d., Gomes, C. S., Bernal, R. T. I., and Malta, D. C. (2022). Violência física contra mulheres perpetrada por parceiro íntimo: análise do viva inquérito 2017. *TEMAS LIVRES • Ciência e saúde coletiva*, 27(10):3993–4002.