

Impacto dos Empreendimentos de Mobilidade Urbana nos Acidentes de Trânsito: Uma Análise Baseada em dados do SIMU

Richard Sousa Antunes, Ana Yoon Faria de Lima

1. Motivação

Este estudo tem como finalidade contribuir para o planejamento estratégico da infraestrutura de mobilidade urbana no Brasil. Empregando dados provenientes da Secretaria Nacional de Mobilidade e Serviços Urbanos (SEMOB) e do Atlas da Violência, busca-se estabelecer uma base analítica para orientar decisões sobre futuras intervenções urbanas. O trabalho focará na avaliação do impacto que os empreendimentos de mobilidade urbana exercem sobre os índices de acidentes de trânsito nas cidades brasileiras. Busca-se determinar quais categorias de empreendimentos têm maior potencial de reduzir esses incidentes, valendo-se de metodologias de ciência de dados. Além disso, o estudo visa estabelecer uma metodologia replicável para o uso dos dados, pavimentando o caminho para análises futuras e aprimoramento contínuo das políticas de mobilidade urbana.

2. Metodologia

A metodologia deste trabalho consiste nos seguintes passos: (1) Análise exploratória dos dados; (2) Limpeza e tratamento de dados; (3) Escolha e implementação de um modelo capaz de relacionar a variável dependente (mortos ou feridos) com o tipo de empreendimento e demais variáveis explicativas; (3) Avaliação do modelo, com testes estatísticos para avaliar a veracidade do modelo e a confiabilidade de seus coeficientes, resultados e extrapolações; (4) Análise de resultados e suas possíveis implicações para políticas de mobilidade urbana.

Detalhes adicionais sobre o modelo e os testes estatísticos implementados são apresentados na Seção 2.1. Em seguida, a Seção 3 detalha como o tratamento de dados e a modelagem foram implementados nos conjuntos de dados estudados. Finalmente, a Seção 4 discute os resultados obtidos. A análise exploratória e os aspectos do desenvolvimento que não foram incluídos na solução final são sumarizados no vídeo que acompanha este relatório. Todos os códigos desenvolvidos podem ser encontrados no repositório do projeto no GitHub: <https://github.com/anayflima/simu-datathon-2024>.

2.1. Modelos

Neste estudo, serão avaliados modelos de Regressão Linear para Dados em Painel (ou Dados Longitudinais), que são particularmente vantajosos para análise de dados que variam tanto no espaço quanto no tempo, permitindo a modelagem simultânea de informações em formato cross-section e em séries temporais (Carneiro, 2021).

O modelo mais básico dessa família, conhecido como Pooled, considera os dados empilhados e não captura variáveis não observadas. O modelo é expresso por:

$$\text{Pooled} : y_{it} = \beta_0 + \beta_1 x_{1it} + \dots + \beta_n x_{nit} + \varepsilon_{it} \quad (1)$$

Onde y_{it} é a variável dependente (neste caso, mortos ou feridos), x_{jit} são as variáveis explicativas selecionadas a partir dos conjuntos de dados, β_j são os coeficientes associados às variáveis explicativas, ε_{it} são os termos de erro e i e t são os índices de municípios e anos, respectivamente.

Este modelo pode ser estendido para dois modelos de dados em painel usados neste trabalho: Efeitos Fixos (EF) e Efeitos Aleatórios (EA), que adicionam termos de efeito fixo α_i e termos de efeito aleatório γ_t , respectivamente:

$$\text{EF} : y_{it} = \alpha_i + \beta_1 x_{1it} + \dots + \beta_n x_{nit} + \varepsilon_{it} \quad (2)$$

$$\text{EA} : y_{it} = \gamma_t + \beta_1 x_{1it} + \dots + \beta_n x_{nit} + \varepsilon_{it} \quad (3)$$

O modelo (2) controla variáveis não observadas que diferem entre as entidades e são constantes no tempo, atribuindo um intercepto para cada entidade. O modelo (3) controla variáveis não observadas que variam no tempo, mas não entre as entidades. Testes estatísticos comparativos, explicados nas próximas subseções, determinam o modelo mais adequado. Adicionalmente, é calculada a medida VIF (Variance Inflation Factor) para detectar possíveis multicolinearidades muito altas com $\text{VIF} > 10$ (Teles, 2023).

2.1.1. Modelo de Efeitos Fixos \times Modelo Pooled

A adequação do Modelo de Efeitos Fixos em relação ao Modelo Pooled é verificada através do Teste de Chow. Este teste opera sob a hipótese nula de que não há diferença significativa entre os interceptos e coeficientes de dois modelos Pooled. Um ponto de corte arbitrário no tempo é selecionado para identificar uma potencial mudança estrutural. Se as análises de regressão indicarem diferenças significativas antes e após este ponto (com um p-valor inferior a 0,05), a hipótese nula é descartada, favorecendo o Modelo de Efeitos Fixos.

2.1.2. Modelo de Efeitos Aleatórios \times Pooled

Para testar o modelo com efeitos aleatórios em relação ao Pooled, é utilizado o **Teste de Breusch-Pagan**, que tem como hipótese nula que o modelo Pooled é homoscedástico. A rejeição desta hipótese indica a existência de heteroscedasticidade¹ não capturada pelo Modelo Pooled, o que justifica a preferência pelo Modelo de Efeitos Aleatórios.

2.1.3. Modelo de Efeitos Fixos \times Efeitos Aleatórios

Caso as hipóteses nulas sejam rejeitadas nos testes anteriores, procede-se com o **Teste de Especificação de Hausman** para comparar a consistência entre o Modelo de Efeitos Fixos e o Modelo de Efeitos Aleatórios. A hipótese nula do teste sugere que o Modelo de Efeitos Aleatórios é mais apropriado. Caso esta hipótese seja rejeitada, o Modelo de Efeitos Fixos é considerado o mais adequado para a análise.

¹Quando a variância dos erros varia entre as observações em um modelo de regressão.

3. Configuração experimental

Para assegurar a reprodutibilidade, o código foi desenvolvido de maneira modular. Funções específicas foram encapsuladas em notebooks Jupyter, que também mostram a aplicação das funções e os resultados derivados. Adicionalmente, as funções também foram isoladas em scripts Python, de forma a facilitar uso futuro.

3.1. Tratamento de dados

Os dados discutidos nesta seção de tratamento são derivados de bases obtidas do site do SIMU: base de empreendimentos, de acidentes, de municípios e de população.

3.1.1. Categorização dos Empreendimentos

Uma das nossas hipóteses é que diferentes tipos de empreendimentos têm impactos distintos no número de mortes e feridos, sendo benéfico classificá-los em categorias, ao invés de tratá-los como um grupo homogêneo. No entanto, os dados de empreendimentos carecem de uma coluna que categorize os empreendimentos em tipos (por exemplo, aqueles relacionados ao transporte público, infraestrutura cicloviária, etc). Por outro lado, a coluna empreendimento está disponível, que contém o nome do empreendimento para cada registro de dados. Um exemplo de valor presente nessa coluna é "CONSTRUCAO DE TREVO DE ACESSO AO MUITIRAO, CICLOVIA DE ACESSO A ZONA INDUSTRIAL".

A partir desta coluna, foram criadas colunas auxiliares indicativas do tipo de empreendimento ao qual cada registro pertence. Inicialmente, realizamos um tratamento textual da coluna de nome, utilizando as bibliotecas `nltk`² e `unidecode` para normalização do texto (remoção de acentos e pontuação e conversão para letras minúsculas) e remoção de palavras de parada ("stop words"), como 'de', 'e', 'a'.

Em seguida, geramos uma nuvem de palavras com as palavras presentes na coluna após o tratamento. O resultado dessa nuvem pode ser visto na Figura 1.



Figura 1: Nuvem de palavras ilustra a frequência dos termos na coluna empreendimento após o tratamento inicial.

Com base nos resultados dessa nuvem de palavras, identificamos algumas categorias de empreendimentos que estão mais presentes nos dados. Desenvolvemos uma função que usa expressões regulares (RegEx) para classificar os registros conforme as categorias identificadas. Fizemos um processo iterativo de categorização e análise das nuvens de

Tabela 1

Categorias de Empreendimentos obtidas a partir dos nomes.

Metrô	Ciclovia	Corredor	Transporte rápido
VLT ³	Praça	Refrota	Pavimentação
Parque	Terminal	Sinalização	Drenagem
Asfalto	Ponte	Calçamento	Recapetamento
Sarjeta	Calçadas	Ônibus	Revitalização
Ferrovia	Aeroporto		

palavras dos registros sem categoria de forma a obter categorias que tenham potencial de informar políticas públicas. As categorias nas quais chegamos por meio desse processo podem ser vistas na Tabela 1.

Para cada uma dessas categorias, criamos uma coluna auxiliar na base de dados que indica se aquele registro pertence a essa categoria. Essas categorias se intersectam em alguns empreendimentos, pois há nomes de empreendimentos que abrangem múltiplas áreas. Também criamos uma coluna agregadora na qual é possível visualizar todas as categorias associadas a um empreendimento.

3.1.2. Tratamento de Dados Ausentes

Os dados de acidentes estão separados por município e por ano. Assim, para possibilitar uma conclusão sobre o impacto dos empreendimentos, iremos agrupá-los por município e por ano. Contudo, identificamos uma quantidade significativa de valores ausentes nessas duas colunas na base de dados de empreendimentos, o que dificultaria a análise de correlação. Assim, procedemos com um tratamento preliminar dos dados para viabilizar agrupamentos mais consistentes e as posteriores análises.

Colunas de Município

Existem diversas colunas que contêm informações municipais na base de empreendimentos. As principais que poderiam definir o município, bem como a quantidade de dados ausentes presentes em cada uma delas, são:

- Código IBGE: 33.554 entradas nulas
- mun_MUNNOME: 41.240 entradas nulas
- mun_MUNNOMEX: 1.207 entradas nulas
- Município: 41.240 entradas nulas
- uf_SIGLA_UF: 1.036 entradas nulas

Como o total de registros na base de empreendimentos é de 71.248 entradas, a proporção de dados ausentes nessas colunas é muito relevante, exceto para a coluna mun_MUNNOMEX. Além disso, a coluna uf_SIGLA_UF, com a Unidade Federativa a qual o município pertence, possui poucos nulos, e observamos que todas as linhas que possuem alguma das outras colunas de município não nulas também possuem a sigla UF preenchida.

Embora a coluna mun_MUNNOMEX apresente menos nulos, ela não é única por Código IBGE, apresentando variações de acentuação e grafia. Além disso, existem municípios homônimos em diferentes estados, o que foi verificado cruzando-se as colunas Código IBGE e mun_MUNNOMEX. No entanto, a combinação das colunas mun_MUNNOMEX e uf_SIGLA_UF é única. Ou seja, não existem municípios de mesmo nome na mesma Unidade Federativa.

²Biblioteca para Processamento de Linguagem Natural.

Queremos maximizar o número de entradas que contém a coluna `Código IBGE`, para que possamos agregar os dados de empreendimentos com os de acidente mantendo a integridade dos dados. Para isso, adotamos a seguinte estratégia:

1. Verificar a consistência dos códigos IBGE na base de dados dos empreendimentos com a base de municípios. Seis ocorrências do código 0 (inexistente na base de municípios) foram corrigidas utilizando as colunas `mun_MUNNOMEX` e `uf_SIGLA_UF` para identificar o código IBGE correto.
2. Tratar a coluna `mun_MUNNOMEX` por meio da remoção de acentos, pontuação e conversão para letras maiúsculas, visando padronização.
3. Utilizar `mun_MUNNOMEX` tratada para preencher os valores de `Código IBGE` ausentes, realizando um merge com a base de municípios e considerando a coluna `uf_SIGLA_UF` para garantir unicidade.
4. Utilizar a coluna `Código IBGE` como chave para integrar a base de empreendimentos com a base de municípios, possibilitando o preenchimento de informações municipais adicionais, tais como coordenadas geográficas, informações de se é capital ou não, etc.

Essas ações resultaram na diminuição significativa dos valores nulos na coluna `Código IBGE`, de 33.554 para 1.126 entradas nulas, e facilitou a integração com a base de municípios e a consequente igual redução de nulos em outras informações municipais relevantes.

Colunas de ano

Identificamos quatro colunas relacionadas a ano na base de empreendimentos:

- ano (ano de registro da obra): 11.583 entradas nulas
- ano_inicio_obra: 23.369 entradas nulas
- ano_fim_obra: 24.704 entradas nulas
- ano_assinatura: 11.583 entradas nulas

Verificamos que as colunas `ano` e `ano_assinatura` são idênticas, tanto nas entradas nulas quanto nos valores não nulos, sendo as colunas que possuem menos nulos entre as colunas de ano. Optamos por seguir com a coluna `ano` no tratamento.

Para preencher as lacunas nas colunas de ano, analisamos a correlação dessas variáveis com outras na base de dados. A coluna `ano` mostrou 92% de correlação com `cod_id_fonte`, sugerindo um padrão evolutivo no tempo. A coluna `ano_inicio_obra` apresentou 93% de correlação com `cod_id_fonte` e 98% com `ano`, enquanto `ano_fim_obra` teve menor correlação (85% com `cod_id_fonte`), refletindo variações na duração das obras.

A coluna `cod_id_fonte` não apresenta entradas nulas, indicando seu potencial como preditor robusto para essas variáveis de ano, dada a sua significativa correlação. Neste contexto, avaliamos alguns algoritmos de Aprendizado de Máquina (Machine Learning - ML) com o intuito de testar a eficácia da `cod_id_fonte` na predição das colunas `ano` e `ano_inicio_obra`. Trata-se de um problema de classificação em ML, não de regressão, já que os anos são valores inteiros discretos. Embora tenha sido explorada a aplicação de um modelo de regressão seguido de arredondamento para o inteiro mais próximo, observou-se uma redução significativa

na acurácia em comparação com a abordagem de classificação direta, uma vez que a acurácia não é o foco principal em problemas de regressão.

Foram utilizados os seguintes algoritmos de classificação, para avaliar como eles performavam: (1) Regressão Logística, (2) K-Vizinhos Mais Próximos⁴ (KNN), (3) Máquinas de Vetores de Suporte⁵ (SVM), (3) Árvore de decisão, e (4) Floresta Aleatória. Os dados foram divididos em conjuntos de treino e teste, e empregou-se as técnicas de Validação Cruzada (Cross-Validation) e de ajuste de hiperparâmetros para identificar o modelo e os hiperparâmetros ótimos dentre os avaliados. O algoritmo KNN, com distância euclidiana, peso uniforme e 106 vizinhos, destacou-se ao alcançar uma acurácia média de 91% na validação cruzada, que já considera a performance do modelo em uma base de testes não utilizada no treinamento.

Por outro lado, a acurácia da `cod_id_fonte` na predição de `ano_inicio_obra` foi substancialmente inferior, atingindo 42% com o modelo SVM, que apresentou o melhor desempenho. Tal discrepância pode ser atribuída à maior quantidade de dados nulos na coluna `ano_inicio_obra` em comparação com `ano`, proporcionando um volume maior de dados para o treinamento do modelo. Adicionalmente, a análise dos dados revelou que o ano de registro (`ano`) tende a ser próximo ao ano de início da obra (`ano_inicio_obra`). Dessa forma, utilizamos a coluna `cod_id_fonte` para prever os valores nulos da coluna `ano` e consideraremos essa coluna como ano dos empreendimentos.

A decisão de considerar o início, e não o término da obra (coluna `ano_fim_obra`), não apenas se beneficia da maior disponibilidade de dados, mas também se alinha à análise de correlação com incidentes, visto que o início de obras de infraestrutura frequentemente acarreta impactos imediatos, tais como interrupções viárias, alterações de rotas e variações de velocidade nas áreas afetadas.

3.2. Agrupamento dos dados

A análise exploratória mostrou que a variável 'população' (`Populacao`) nos dados de empreendimentos deriva da população municipal no ano de registro da obra. Após o preenchimento dos valores nulos nas colunas `Código IBGE` e `ano`, a integração com a base de população reduziu os nulos em `Populacao` de 41.240 para 1.126.

Posteriormente, procedeu-se ao agrupamento dos dados por município e ano. Desenvolveu-se uma função específica para processar os dados de empreendimentos, retornando, para cada par município-ano, os valores totais de investimento e os montantes investidos em cada categoria de empreendimento (descritas na Seção 3.1.1. Adicionalmente, a função calcula o total da população beneficiada (coluna `pop_beneficiada`), o número de empreendimentos agregados e o total de investimento por tipo de programa (coluna `programa` da base de dados original).

Com os dados de empreendimentos organizados por município e ano, eles foram combinados com as bases do

⁴K-Nearest Neighbors, em inglês.

⁵Support Vector Machines, em inglês.

Atlas de Violência, que contém informações sobre mortes e feridos. Esta integração resultou em uma base de dados consolidada, abrangendo tanto os detalhes dos empreendimentos, categorizados por tipo e programa, quanto as estatísticas de acidentes.

Com a base de dados unificada, é possível avançar para a etapa subsequente, que consiste na análise de correlação entre os fatores de empreendimentos e os índices de acidentes, visando identificar padrões e possíveis relações.

3.3. Modelo

Após o tratamento dos dados, os procedimentos descritos nesta seção foram realizados duas vezes, de forma independente, considerando y como o número total de mortos e, posteriormente, como o número total de feridos. Ambas as variáveis apresentaram resultados análogos nos testes estatísticos e na seleção do modelo final.

Inicialmente, o modelo simples Pooled 1 foi implementado em conjunto com o teste de Chow. Obteve-se um p -valor < 0.05 , levando à rejeição da hipótese de que o modelo Pooled seria superior ao de Efeitos Fixos. Em seguida, o modelo Pooled 1 foi implementado em conjunto com o teste de Breusch - Pagan. Obteve-se p -valor < 0.05 , rejeitando-se a hipótese de que o modelo Pooled seria superior ao de Efeitos Aleatórios.

Com base nos resultados dos dois testes mencionados, descartou-se a validade do modelo Pooled, indicando a necessidade de utilizar modelos específicos para dados em painel, tais como os modelos de Efeitos Fixos (2) e de Efeitos Aleatórios (3). Assim, o Teste de Hausman foi implementado para escolher entre esses dois modelos. Obteve-se um p -valor > 0.05 , não rejeitando a hipótese nula, ou seja, o **Modelo de Regressão para Dados em Painel com Efeitos Aleatórios** (3) foi escolhido como o modelo ideal para ambos os casos de $y = \text{total mortes}$ e $y = \text{total feridos}$.

Os coeficientes estimados pelo modelo de Efeitos Aleatórios, escolhido como melhor, para o número total de mortes estão apresentados na Tabela 2. Os resultados para o número total de feridos obtidos por esse modelo foram similares aos do número de mortes, com variações em algumas categorias, e pode ser visto no repositório do GitHub. Adicionalmente, calculou-se o fator VIF, que se manteve abaixo de 5 para todas as variáveis, indicando a ausência de multicolinearidade significativa. Coeficientes negativos sugerem um impacto negativo da respectiva categoria no total de mortes. Ou seja, municípios que investiram nessa categoria observaram uma tendência de redução no número total de mortes ao longo do tempo. Por outro lado, coeficientes positivos indicam uma relação positiva com o número de mortes. Categorias que não apresentaram significância estatística (com p -valor superior a 0.05) e, portanto, não permitiram uma correlação clara com o número de mortes, não foram incluídas na tabela.

O resultado desses coeficientes e possíveis implicações para políticas públicas estão descritos na Seção 4.

Tabela 2

Coeficientes válidos obtidos a partir do modelo de Efeitos Aleatórios para o total de mortes.

Coeficientes significativos (p -valor $\ll 0.05$)			
Categoria	Coef.	Categoria	Coef.
Invest. p/ capita	$-1.75 \cdot 10^{-2}$	Pavimentacao	$4.39 \cdot 10^{-7}$
Refrota	$-3.34 \cdot 10^{-6}$	aeroporto	$2.09 \cdot 10^{-7}$
Ciclovias	$-4.82 \cdot 10^{-7}$	Ferrovia	$1.64 \cdot 10^{-7}$
Drenagem	$-1.93 \cdot 10^{-7}$	Rápido	$1.59 \cdot 10^{-7}$
Metrô	$-1.11 \cdot 10^{-8}$	Ônibus	$1.53 \cdot 10^{-7}$
Sarjeta	$1.67 \cdot 10^{-6}$	Calçadas	$5.88 \cdot 10^{-8}$
Terminal	$6.07 \cdot 10^{-7}$	Corredor	$3.48 \cdot 10^{-8}$
Ponte	$4.75 \cdot 10^{-7}$	-	-

4. Principais resultados

Conforme explorado nas Seções 2.1 e 3.3, o modelo empregado foi capaz de isolar de maneira eficaz o impacto de diferentes categorias de empreendimentos ao longo do tempo, controlando o efeito intrínseco a cada município. Nessa seção, discutiremos os principais resultados obtidos do modelo. Adicionalmente, apresentaremos um mapa interativo dos municípios brasileiros que foi desenvolvido a fim de facilitar a visualização dos dados e promover análises futuras. Este mapa integra, em uma única plataforma, os índices de acidentes e os valores de investimentos em mobilidade urbana.

4.1. Resultados do modelo

A análise dos coeficientes do modelo revelou que diferentes categorias de empreendimentos têm efeitos variados sobre os índices de acidentes. Um coeficiente de valor x indica que um investimento de 1 real nessa categoria está associado a uma variação de x no número de óbitos. Isso sugere que um investimento de $1/x$ reais nessa categoria está associado a um aumento ou diminuição de 1 morte. Embora isso não estabeleça causalidade direta, essa interpretação fornece uma direção para a compreensão dos dados e da magnitude do impacto estimado pelo modelo.

Observou-se que empreendimentos ligados a investimentos rodoviários e aumento da velocidade viária, como pavimentação e expansão de corredores de ônibus, estão associados ao aumento de fatalidades. Isso pode ser atribuído ao comportamento de motoristas que, em vias melhor pavimentadas, tendem a aumentar a velocidade, elevando o risco de acidentes fatais. Embora isso não deva implicar na suspensão desse tipo de investimentos dentro da infraestrutura urbana, ressalta-se a necessidade de medidas mitigadoras, como a fiscalização de velocidade por meio de radares, por exemplo.

De forma surpreendente, o investimento em calçadas mostrou-se positivamente relacionado ao número de mortes. Como consideramos o início da obra como ano do empreendimento, uma hipótese para essa observação é que, durante a restauração das calçadas, pedestres podem ser forçados a caminhar pelas vias, aumentando temporariamente o risco de fatalidades. No entanto, as calçadas possuem uma relação

negativa com o número de feridos, sugerindo que a restauração e a construção de novas calçadas melhoram a segurança a longo prazo. Portanto, recomenda-se a provisão de rotas alternativas seguras para pedestres durante essas obras de infraestrutura.

Por outro lado, percebe-se que o investimento em algumas modalidades de transporte alternativas ao rodoviário, como ciclovias e metrô, demonstraram uma associação negativa com o número de mortes. A análise exploratória dos dados de acidentes nos mostraram que 6.7% das mortes são de ciclistas, porcentagem desproporcional à parcela de viagens feitas de bicicleta no país, o que poderia explicar o impacto negativo significativo de ciclovias no número de mortes. Além disso, a implementação de ciclovias tende a reduzir a velocidade do tráfego, o que geralmente reduz o número de fatalidades. O metrô, ao oferecer uma alternativa eficiente ao transporte rodoviário, com menos incidentes, também apresenta uma relação negativa.

Além disso, melhorias na qualidade do transporte público, representadas na categoria 'Refrota', refletiram em uma diminuição nas taxas de mortalidade. Investimentos em sistemas de drenagem também mostram uma relação negativa com as fatalidades, uma vez que aprimoram as condições das vias ao reduzir inundações e pistas escorregadias, mitigando assim o risco de acidentes de trânsito.

Além disso, o modelo indicou a existência de um coeficiente negativo associado ao investimento per capita. Isso implica que incrementos nos investimentos em mobilidade urbana, considerando a proporção por habitante, estão vinculados a uma diminuição no número total de mortes. Em outras palavras, um incremento nos recursos destinados à infraestrutura de mobilidade urbana por indivíduo está associado a uma tendência de queda no número de acidentes fatais.

Em suma, a análise dos coeficientes fornece insights valiosos para órgãos públicos na avaliação do impacto de empreendimentos urbanos. Esses dados podem orientar decisões estratégicas que visem a combinação de diferentes tipos de empreendimentos para maximizar a segurança viária e minimizar o número de mortes e feridos. A implementação de políticas públicas baseadas em evidências pode, portanto, ser um caminho promissor para a melhoria contínua da mobilidade urbana e da qualidade de vida nas cidades.

4.2. Visualização Interativa de dados municipais

Com os dados tratados e organizados por município, desenvolvemos um mapa interativo, acessível em <https://mapa-empreendimentos-acidentes.fly.dev/>, que proporciona uma visualização integrada das informações. O mapa, ilustrado na Figura 2, permite aos usuários selecionar variáveis de investimento — detalhadas por categoria e programa — e variáveis de acidentes, incluindo mortes e feridos discriminados por meio de transporte.

Além disso, o mapa oferece a opção de visualizar os dados de investimento e acidentes per capita, permitindo isolar o tamanho populacional dos municípios. O raio dos pontos representa proporcionalmente o valor da variável

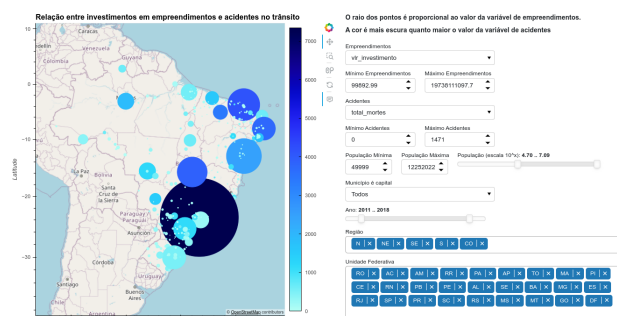


Figura 2: Ilustração do Mapa interativo com alguns filtros.

de empreendimentos, enquanto a cor indica a magnitude da variável de acidentes — pontos mais escuros denotam valores mais altos.

Os dados abrangem todos os anos com registros de acidentes disponíveis, e os usuários podem filtrar o período de interesse. Adicionalmente, o mapa permite filtrar municípios por tamanho populacional, diferenciar entre capitais e não capitais, e selecionar regiões ou unidades federativas específicas para visualização.

Uma aplicação prática do mapa é a identificação rápida e visual de cidades para análises detalhadas. Por exemplo, é possível discernir quais municípios se destacam em termos de investimento e incidência de acidentes. O mapa também facilita a comparação entre cidades com valores de investimentos total similares, mas com discrepâncias significativas nos números de mortes ou feridos, suscitando questões sobre os fatores que influenciam tais diferenças.

Um caso de estudo comparativo foi realizado entre duas cidades do Nordeste — Fortaleza e Salvador — que, apesar de apresentarem valores próximos de investimento total, exibiram diferenças marcantes no número de mortes. A análise detalhada revelou que Fortaleza registrou mais mortes de pedestres e automóveis, enquanto Salvador teve um número maior de mortes entre ocupantes de ônibus. Ao examinar a distribuição dos investimentos por categoria, observou-se que Salvador investe proporcionalmente mais em renovação de frota, corredores de ônibus, calçadas e transporte coletivo, enquanto Fortaleza concentra seus recursos em metrô.

Este mapa interativo serve como uma ferramenta valiosa para cientistas de dados na condução de análises exploratórias e na determinação de focos de investigação. A funcionalidade de visualização interativa também torna a ferramenta acessível a indivíduos não técnicos e agentes públicos interessados em obter uma visão geral dos investimentos e dos acidentes de trânsito em municípios específicos.

Referências

- Carneiro, D.R.F., 2021. Dados em painel. <https://drive.google.com/file/d/1bxRmw1uTG0XNhnS0mkpRA20gXK382ZC8/view>. Curso em slides.
- Teles, T., 2023. Vif (variance inflation factor). <https://www.linkedin.com/pulse/vifvariance-inflation-factor-thiago-teles/>. Artigo no LinkedIn.