

Datathon de Mobilidade Urbana 2024

Proponente: Vinícius da Fonseca Vieira

Email: vfvieira@gmail.com

CPF: 064666666-54

Video no youtube: <https://youtu.be/2RUnJYB4eGk>

Dados: <https://encurtador.com.br/vgJBS>

Github do projeto: <https://github.com/vfvieira/datathon2024/>

Junho de 2024

Motivação da proposta

Este documento apresenta a solução desenvolvida como resposta ao **Datathon de Mobilidade Urbana 2024**, promovido pela Secretaria Nacional de Mobilidade Urbana (SEMOB) do Ministério das Cidades (MCID), em parceria com a Plataforma de Ciência de Dados aplicada à Saúde (PCDaS), iniciativa do Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (Icict) da Fundação Oswaldo Cruz (Fiocruz). A proposta apresentada visa responder ao **Desafio 1**: Análise dos empreendimentos em mobilidade urbana da Carteira de Empreendimentos da MCID/SEMOB e sua correlação no território com os indicadores de mortos e feridos no trânsito mantidos pelo Atlas da Violência, considerando o tipo de programa do empreendimento.

Assim, tomando como base a **Carteira de Empreendimentos da MCID/SEMOB** e os **indicadores de mortos e feridos no trânsito mantidos pelo Atlas da Violência**, toda a solução desenvolvida foi motivada pela investigação da hipótese de que os indicadores de mortos e feridos nos municípios são diretamente impactados pelos empreendimentos realizados pelo MCID. Nesse sentido, algumas questões foram formuladas para conduzir os experimentos apresentados nesta proposta: “**Questão 1**: Como se caracteriza, ao longo do tempo, a relação entre os empreendimentos em mobilidade urbana realizados do MCID e os principais indicadores de mortos e feridos no trânsito?”; “**Questão 2**: Como essa relação varia em razão do programa ao qual os empreendimentos estão associados?”; **Questão 3**: Como os dados sobre os empreendimentos de mobilidade podem ser utilizados para propor a redução em indicadores de mortos e feridos no trânsito?”.

Uma breve investigação sobre os dados

Para a condução desta proposta, primeiramente foi fundamental estudar as principais variáveis contidas nas duas principais bases de dados consideradas (Carteira de Empreendimentos e Indicadores de Violência). O objetivo desta análise é conhecer a natureza dos dados, com foco na descoberta daqueles com maior potencial de responder às questões propostas. Da Carteira de Empreendimentos, cuja cada linha representa um empreendimento distinto, foram selecionadas, inicialmente, as colunas: **programa** (nome do programa relacionado à obra, texto); **Código IBGE** (código do IBGE dos municípios onde a obra está localizada, valor inteiro); **pop_beneficiada** (população beneficiada pela obra, valor inteiro) **vlr_investimento** (valor total do investimento, valor real); **ano_fim_obra** (ano previsto de conclusão da obra, valor inteiro). Da base de dados de mortos e feridos, cuja cada linha apresenta os indicadores de mortos e feridos em cada município Brasileiro em cada ano entre 2010 e 2019, foram selecionadas as seguintes colunas: **Código IBGE** (código do IBGE dos municípios onde a obra está localizada, valor inteiro); **ano** (ano de registro dos indicadores no município, valor inteiro); **mortes** (total de mortes no município em um determinado ano); **total_feridos** (total de feridos no município em um determinado ano). Devido ao alto índice de valores faltantes na coluna **População** da Carteira de Empreendimentos e um alto índice de valores com erros nas colunas **taxa_mun_mortes** e **taxa_mun_feridos** na base de acidentes, informações sobre a população dos municípios foram obtidas a partir de estimativas populacionais descritas no Sistema IBGE de Recuperação Automática (SIDRA)¹, com o objetivo de encontrar valores proporcionais às populações das cidades para as propriedades estudadas.

Ainda buscando delimitar um escopo do conjunto de dados a serem considerados na solução, foi realizada uma contagem do número de empreendimentos, relacionados a cada programa, finalizados em cada ano. Linhas com valor ausente nas colunas **ano_fim_obra** ou **programa** foram desconsideradas. A Tabela 1 ilustra o resultado encontrado nessa contagem.

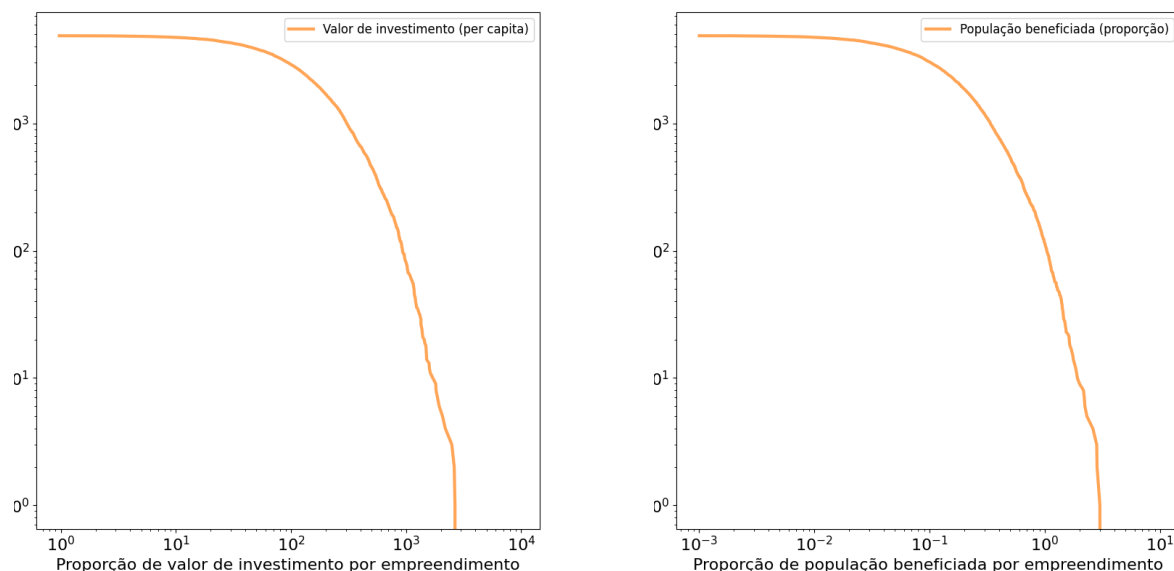
¹ <https://sidra.ibge.gov.br/pesquisa/estimapop/tabelas>

Ano	AVANÇAR PÚBLICO	EMENDAS	PAC FIN	PAC OGU	PRÓ COMUNIDADE	SETOR PRIVADO
2000					2	
2002					1	
2003					1	
2008						1
2012			5			1
2013			13			
2014			32			
2015		22916	37	2		
2016		1111	50	1		
2017		1748	66	4	1	1
2018		2398	53	4		51
2019		3046	103	4		42
2020	6	3467	88	6		7
2021	41	3533	46	5		
2022	43	2943	36	3		10
2023	13	958	3	9		

Tabela 1: Número de empreendimentos por programa por ano na Carteira de Empreendimentos de Mobilidade Urbana.

Um primeiro recorte foi necessário no estudo para o intervalo de 2010 a 2019, que é o intervalo contido na base de dados de Mortos e Feridos (Atlas de Violência). Adicionalmente, foram descartados os dados relativos a programas em anos com baixa frequência, restando, nas análises seguintes, apenas informações relacionadas a programas e anos com frequência relevante (valores em destaque na Tabela 1).

Com o objetivo de melhor compreender os dados analisados, foram gerados gráficos de Função de Distribuição Cumulativa Complementar (Complementary Cumulative Distribution Function - CCDF²), para as propriedades: valor de investimento e população beneficiada (proporcionais à população do município, por empreendimento) e mortes e feridos no trânsito (média por município, por 100k habitantes). Esses gráficos são apresentados na Figura 1.



² <https://www.ni.com/docs/en-US/bundle/rfm-x-specan/page/ccdf.html>

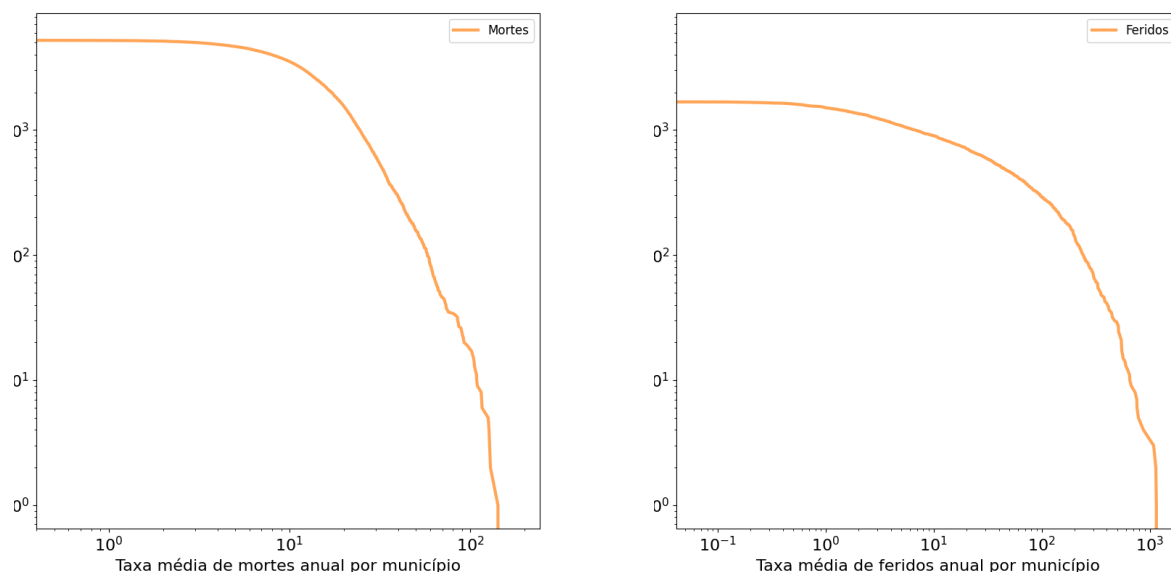
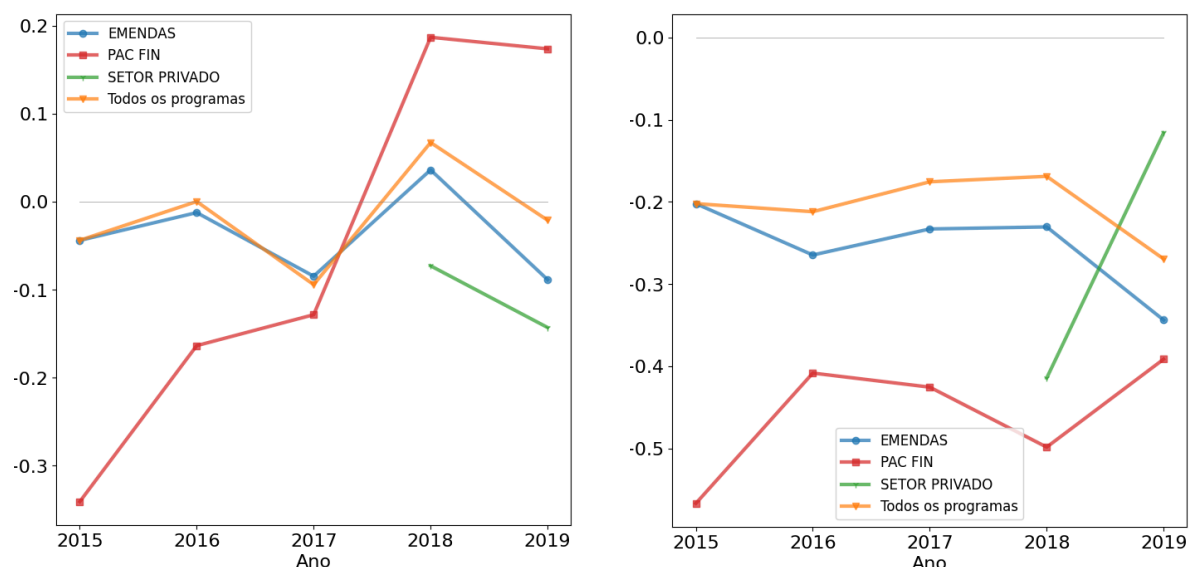


Figura 1: Função de Distribuição Cumulativa Complementar das variáveis: valor de investimento (*per capita*), população beneficiada (*per capita*), mortes e feridos.

Uma investigação visual das funções de distribuição das variáveis sugeriu uma alta redundância entre as propriedades *vlr_investimento* e *pop_beneficiada*, o que foi confirmado pela alta correlação de Spearman³ de suas respectivas colunas na base de dados original (**=0.97**). Optou-se por abandonar a propriedade *pop_beneficiada* nas análises seguintes, considerando que não é capaz de trazer informação que já não esteja contida na propriedade *vlr_investimento*.

Correlação estática: relacionando propriedades em uma mesma época

Visando responder às questões propostas, a primeira análise foi feita sob uma perspectiva estática com o objetivo de correlacionar o **valor de investimento** (calculado *per capita*) em cada cidade com os indicadores de **mortos** e **feridos** em cada ano. Para isso, a Carteira de Empreendimentos foi percorrida e o valor *per capita* do investimento de cada empreendimento, acumulado para cada cidade em cada ano, foi correlacionado com os indicadores de mortos e feridos naquela cidade no mesmo ano, considerando uma estratificação por programa. Foi utilizada a correlação de **Spearman**. Os resultados são apresentados na Figura 2.



³ <https://statistics.laerd.com/statistical-guides/spearman-rank-order-correlation-statistical-guide.php>

Figura 2: Correlação entre o valor de investimento em cada município e indicadores de mortes e feridos em cada ano do intervalo entre 2015 e 2019.

Como apresentado pela Figura 2, não é possível observar uma correlação significativa entre o valor de investimento e o número de mortes e feridos quando consideramos uma correlação de seus valores em cada um dos anos. Há uma correlação moderada entre o valor de investimento e o número de feridos no ano de 2015 para o programa PAC FIN (mesmo com uma frequência significativamente inferior a empreendimentos provenientes de EMENDAS). É interessante notar também que a curva de correlação quando todos os programas são considerados é dominada por programas do tipo EMENDAS, que apresentam volume significativamente maior.

Correlação dinâmica: relacionando propriedades em épocas diferentes

Os resultados apresentados pelo estudo de correlação estática levantam a necessidade de que novas visões sobre a relação entre as propriedades de interesse sejam exploradas. Nesse sentido, um experimento sobre a correlação dinâmica entre as propriedades foi conduzido com o objetivo de verificar se o investimento em empreendimentos de mobilidade impacta os indicadores de mortos e feridos no futuro (e quando).

Para esse estudo, um primeiro passo foi feito com a criação de séries temporais para as variáveis relacionadas às propriedades de interesse (vlr_investimento, mortes e feridos). Cada série temporal armazena, por ano, a média dos valores da respectiva propriedade. A partir da definição das séries temporais com as médias de valor de investimento, mortes e feridos em cada um dos anos, foi realizado um estudo de Correlação Cruzada com Atraso de Tempo⁴ (Time-Lagged Cross Correlation - TLCC), onde os valores de uma série temporal são correlacionados com os valores de outra série temporal, mas com diferentes atrasos (*lag*). Assim, podemos investigar, por exemplo, qual o impacto de um investimento aplicado, por exemplo, no ano de 2016 aos indicadores de mortes e feridos nos anos seguintes (2017, 2018, 2019). Neste estudo e nos próximos, abandonamos também o programa SETOR PRIVADO, que apenas apresenta registros em dois anos, impedindo a avaliação dinâmica de seus empreendimentos. A Figura 3 apresenta os resultados de TLCC entre o valor de investimento e os mortos e feridos nos anos seguintes.

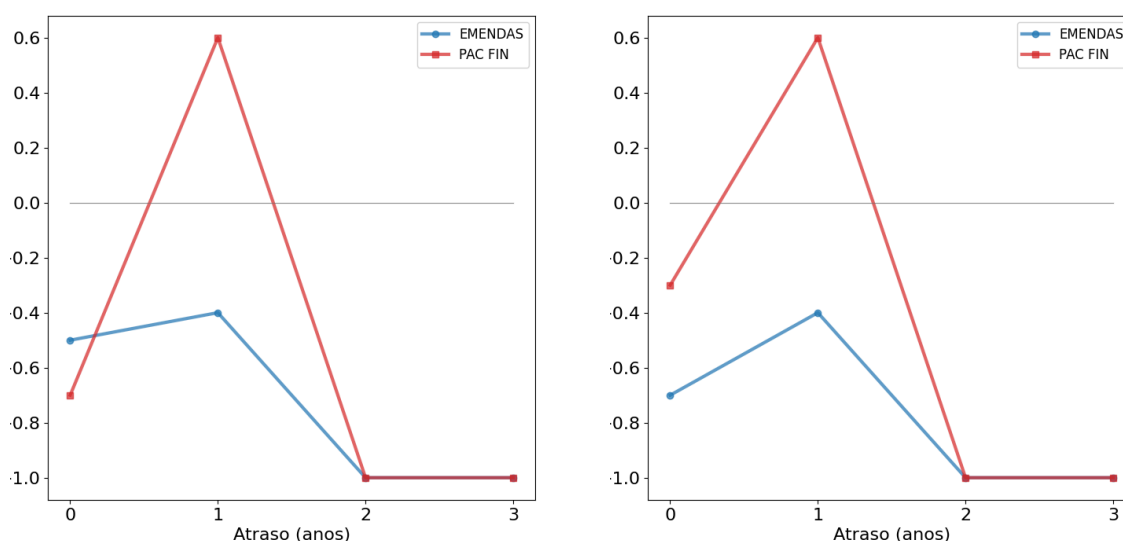


Figura 3: TLCC entre o valor de investimento e indicadores de mortes e feridos agregados em cada em cada ano com diferentes atrasos (em anos) do intervalo entre 2015 e 2019.

O resultado da Figura 3 nos mostra que mesmo que a correlação entre as propriedades seja apenas moderada quando as analisamos nos municípios ano-a-ano, há uma correlação significativa entre as propriedades quando são analisadas de forma agregada ao longo do tempo. O resultado sugere que um empreendimento de mobilidade urbana **não tem impacto imediato** nos indicadores de mortos e feridos, mas isso ocorre após um intervalo de **2 a 3 anos**. Isso reforça o caráter estratégico e de mais longo prazo da aplicação de recursos públicos para a melhoria da qualidade de vida de uma população. Esse impacto não tem

⁴ <https://doi.org/10.1016/j.physleta.2014.12.036>

distinção quanto ao programa ao qual está relacionado o empreendimento.

Variação de uma propriedade sob a perspectiva de outra

Continuando na avaliação da relação entre as propriedades da Carteira de Empreendimentos de Mobilidade Urbana e de Mortes e Feridos (Atlas de Violência), foi realizado um experimento visando compreender como é o comportamento de uma propriedade sob a perspectiva de outra. Duas perspectivas foram consideradas: a perspectiva da propriedade valor de investimento e a perspectiva das propriedades mortes e feridos.

Avaliação sob a perspectiva de valor de investimento

O objetivo desta análise, cujo resultado é apresentado na Figura 4, é avaliar como as propriedades **mortes** e **feridos** estão distribuídas por municípios com diferentes escalas da propriedade **valor de investimento**. Para isso, foi calculado, por município, o valor de investimento acumulado, estratificado por programa. Em seguida, foram calculados os quartis (Q1, Q2, Q3 e Q4) dos valores. Para os municípios em cada um desses quartis (eixo-x da Figura 4) foram calculadas as variações de mortes e feridos no intervalo entre 2015 e 2019 (eixo-y) para os municípios com redução de mortes e feridos.

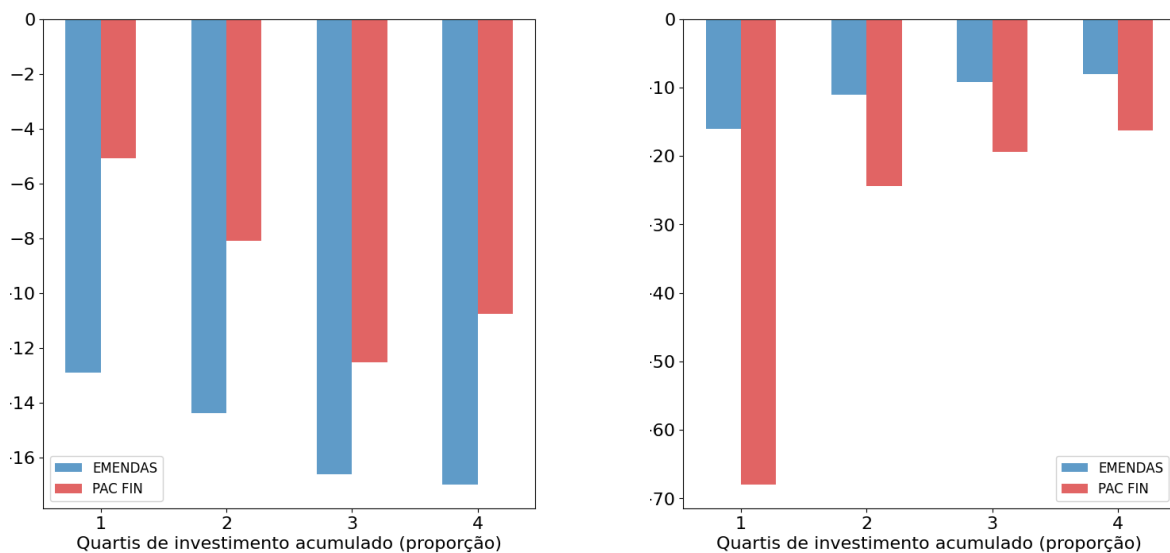


Figura 4: Avaliação da variação dos indicadores de mortes e feridos sob a perspectiva de valores de investimento acumulados por município em diferentes quartis.

O resultado da avaliação sobre mortes (Figura 4 à esquerda) nos mostra que os quartis superiores apresentaram variações significativamente mais importantes do que os quartis inferiores. Isso reforça a verificação da hipótese de que um maior investimento *per capita* em empreendimentos de mobilidade está relacionado à redução de mortes em uma janela de tempo de médio e longo prazo. Mais ainda, é possível perceber que os programas do tipo EMENDA, que possuem um número muito superior de registros, trazem impacto significativamente maior do que os programas do tipo PAC FIN, especialmente em quartis inferiores. Um resultado muito diferente pode ser visto para os resultados sobre feridos (Figura 4 à direita). Os quartis inferiores apresentam variações significativamente mais importantes do que os quartis da direita. Além disso, investimentos do tipo PAC FIN trazem um impacto muito superior aos do tipo EMENDAS. Isso pode ocorrer por alguns motivos. Um deles é o alto número de valores ausentes de registros de feridos na base de dados. Outro motivo é a possibilidade de que obras de menor impacto, portanto, de menor valor de investimento, sejam, de fato mais eficientes na redução de acidentes menos graves, que não acarretam em vítimas fatais.

Avaliação sob a perspectiva de mortes e feridos

Quando a perspectiva de mortes e feridos é considerada, o objetivo é avaliar como a propriedade **valor de investimento** está distribuída por municípios com diferentes escalas das propriedades **mortes** e **feridos**.

Para que o resultado desta análise, cujo resultado é apresentado na Figura 5, foi calculado, por município,

a variação de mortes e feridos acumulados, estratificados por programa. Em seguida, foram calculados os quartis (Q1, Q2, Q3 e Q4) dos valores. Para os municípios em cada um desses quartis (eixo-x da Figura 5) foram calculadas os valores de investimento acumulados no intervalo entre 2015 e 2019 (eixo-y).

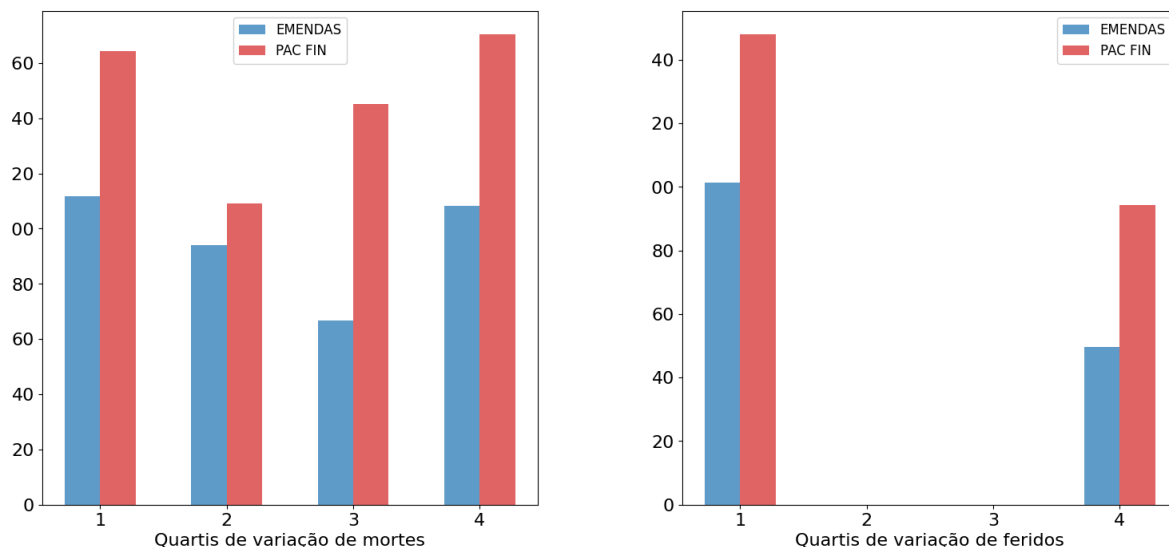


Figura 5: Avaliação de valores de investimento acumulados por município sob a perspectiva da variação dos indicadores de mortes e feridos em diferentes quartis.

Quando olhamos os resultados sob a perspectiva da variação de mortes (Figura 5 à esquerda), os municípios no quartil 1 e 4 (ou seja, os municípios com maior e menor redução de mortes, respectivamente) são aqueles com maior valor de investimento acumulado. É possível que as maiores reduções de mortes, observadas em municípios do Q1, estejam relacionadas às maiores taxas de investimento em empreendimentos nesses municípios. É possível também que municípios em outros quartis, com menores reduções em mortes, sejam aqueles que já possuem valores baixos de mortes no início do intervalo, eliminando as relações entre as variáveis nesse recorte. Assim, embora seja possível elaborar algumas conjecturas sobre esses resultados, não é possível extrair conclusões definitivas.

Sob a perspectiva dos municípios organizados pelos números de feridos, chama a atenção a ausência de valores nos quartis Q2 e Q3. Isso ocorre devido a valores nulos de feridos registrados em um elevado número de municípios, que pode ser ocasionado por uma subnotificação de vítimas de acidentes não-fatais. Mesmo assim, quando observamos os quartis Q1 e Q4, notamos que a hipótese é verificada e que municípios com maiores reduções em feridos apresentam maiores valores de investimento.

“E se...”: Explorando cenários alternativos

Após analisar os indicativos de mortes e feridos nos municípios que receberam empreendimentos de mobilidade urbana, este estudo propõe uma visão sobre os municípios que **não foram contemplados** com tais empreendimentos. O objetivo é investigar a projeção de mortes e feridos em municípios que não foram contemplados com empreendimentos de mobilidade urbana com base em informações sobre municípios que foram contemplados.

Para isso, foram realizados os seguintes passos:

- 1) Listar todos os municípios da base do Atlas de Violência que **não** estão contemplados na carteira de empreendimentos.
- 2) Para cada um desses municípios, descobrir, na carteira de empreendimentos, aquele que tem a taxa de mortes (ou feridos) mais parecida com a sua em 2015 (ano do início do intervalo). Esse será considerado seu **município mais próximo**.
- 3) Observar a taxa de variação de mortes (ou feridos) do seu município mais próximo entre 2015 e 2019.
- 4) Aplicar ao município não contemplado a taxa de variação de mortes (ou feridos) de seu município mais próximo.
- 5) Observar o número absoluto de mortes (ou feridos) que o município não contemplado teria em 2019 nesse cenário projetado.
- 6) Comparar o número de mortes (ou feridos) projetado com o valor real registrado na base de indicadores. Essa será a diferença entre o cenário projetado e o cenário real.

Por exemplo:

- O município de código **3303807** tem 40478 habitantes e não foi contemplado na carteira de empreendimentos.
- A taxa de mortes de **3303807** em 2015 é **93.8781** e em 2019 é **108.8845** por 100k habitantes.
- O município com taxa mais próxima é o de código **5300108**, com taxa 99.6970 por 100k habitantes em 2015.
- O município mais próximo teve uma **redução de 12.3085 mortes por 100k habitantes** em 2019, passando para 87.3886 mortes por 100k habitantes.
- Quando aplicamos essa redução ao município não contemplado, sua taxa projetada em 2019 passa a ser 96.5760.
- Ou seja, nesse cenário projetado, **4.9822 mortes teriam sido evitadas apenas em 2019 no município 3303807**.

A base de indicadores de acidentes do Atlas de Violência apresenta **5570** municípios. Desses, **4898 estão contemplados** na carteira de empreendimentos da SEMOB. Ou seja, há **672 municípios que não estão contemplados**. Desses, **433** apresentam registros de morte em 2015 e **126** apresentam registros de feridos em 2015. Quando aplicamos o método de projeção a esses municípios, temos:

Mortes evitadas

134 mortes seriam evitadas nos **433 municípios apenas em 2019** (a um custo de R\$41,58* por pessoa em empreendimentos de mobilidade urbana)

Feridos evitados

57 feridos seriam evitados nos **126 municípios apenas em 2019** (a um custo de R\$57,31* por pessoa em empreendimentos de mobilidade urbana)

*Tomando como base o valor de investimentos dos empreendimentos dos municípios mais próximos no intervalo entre 2015 e 2019.

Considerações finais

O desenvolvimento desta solução como resposta ao Datathon de Mobilidade Urbana 2024 trouxe vários desafios ao longo do caminho, mas foi possível encontrar correlações entre propriedades na Carteira de Empreendimentos e Indicadores de Mortes e Feridos (Atlas de Violência) em diferentes perspectivas. Tomando como base a propriedade que descreve os valores de investimento dos empreendimentos, que foi considerada como aquela que apresentava maior potencial de correlação com os indicadores de mortos e feridos, percebemos que empreendimentos de mobilidade urbana não tem correlação imediata com indicadores de mortes e feridos, mas quando consideramos um atraso de 2 a 3 anos, os empreendimentos, representados pelo seu valor de investimento, têm impacto significativo sobre acidentes no trânsito. Mesmo quando consideramos as propriedades de uma base de dados sob a perspectiva de propriedades de outra, percebemos que há relações que existem e podem ser quantificadas com a metodologia proposta.

É importante lembrar que cidades são sistemas complexos e indicadores de violência no trânsito, como mortes e feridos, não estão relacionados apenas a empreendimentos de mobilidade em um intervalo de 5 anos, mas, pelo menos, em toda uma política pública de construção de infraestrutura de mobilidade por um longo prazo. Do ponto de vista de análise de dados, é importante notar que quanto mais fina a granularidade dos dados, mais conclusivas podem ser as observações realizadas. Por exemplo, a geolocalização dos empreendimentos e dos acidentes em nível de setores censitários poderia permitir avaliar o impacto dos empreendimentos em regiões das cidades com diferentes indicadores socioeconômicos. É importante notar também, ainda sob o ponto de vista de análise de dados, que mortes e feridos são propriedades de ocorrência rara, o que impõe um desafio na tarefa de identificação de correlações. Mesmo assim, um estudo sobre os municípios que não foram contemplados com empreendimentos de mobilidade urbana no intervalo estudado, nos ajuda a ter uma noção mais concreta das vidas que poderiam ser salvas, caso tais empreendimentos tivessem sido implementados. Assim, é possível argumentar que o investimento em empreendimentos de mobilidade urbana é fundamental, não só para a melhoria da qualidade de vida humana nos centros urbanos do o ponto de vista de locomoção e garantia do direito à cidade, mas do ponto de vista de indicadores de saúde, elevando a dignidade das pessoas na nossa sociedade.