



**UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
CURSO DE ENGENHARIA CIVIL**

FRANCISCO NILSO DE BRITO FILHO

**ANÁLISE DO IMPACTO DE FAIXAS DE CIRCULAÇÃO EXCLUSIVA AO
TRANSPORTE COLETIVO EM FORTALEZA USANDO *BIG DATA***

**FORTALEZA
Abril/2021**

FRANCISCO NILSO DE BRITO FILHO

**ANÁLISE DO IMPACTO DE FAIXAS DE CIRCULAÇÃO EXCLUSIVA AO
TRANSPORTE COLETIVO EM FORTALEZA USANDO *BIG DATA***

Monografia apresentada ao curso de Engenharia Civil do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial para obtenção do título de Engenheiro Civil.

Orientador: Prof. Ph.D. Francisco Moraes de Oliveira Neto

FORTALEZA
2021

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

B875a Brito Filho, Francisco Nilso de.

Análise do impacto de faixas de circulação exclusiva ao transporte coletivo em Fortaleza usando big data / Francisco Nilso de Brito Filho. – 2021.
82 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Tecnologia, Curso de Engenharia Civil, Fortaleza, 2021.

Orientação: Prof. Dr. Francisco Moraes de Oliveira Neto.

1. Big Data . 2. Transporte Coletivo . 3. Faixas Exclusivas . 4. Inferência Causal . 5. Diferença nas Diferenças. I. Título.

CDD 620

FRANCISCO NILSO DE BRITO FILHO

ANÁLISE DO IMPACTO DE FAIXAS DE CIRCULAÇÃO EXCLUSIVA AO
TRANSPORTE COLETIVO EM FORTALEZA USANDO *BIG DATA*

Monografia apresentada ao curso de Engenharia Civil do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial para obtenção do título de Engenheiro Civil.

Aprovada em 15/04/2021

BANCA EXAMINADORA

Prof. Ph.D. Francisco Moraes de Oliveira Neto (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Ph.D. Manoel Mendonça de Castro Neto
Universidade Federal do Ceará (UFC)

Prof. MSc. Nelson de Oliveira Quesado Filho
Centro Universitário Christus

AGRADECIMENTOS

À minha mãe, Iracilda, por sempre ter me incentivado e dado apoio incondicional para que eu pudesse ter chegado até aqui. Amo muito você.

À minha família, ao meu pai, Nilso, meus irmãos Laion e Zizi, e minhas irmãs Maria e Kátia, às minhas tias e tios, primas e primos e todos que contribuíram de alguma forma para minha formação pessoal. Amo demais todos vocês.

Ao meu orientador, professor Moraes, por quem tenho grande admiração e só tenho a agradecer pela orientação e pela disponibilidade de contribuir com este trabalho. Além disso, gostaria de agradecer-lo pelas conversas, conselhos e por sempre ter acreditado no meu potencial.

Ao meu irmão Laion, por ter-me oferecido moradia e acolhimento, me permitido ir as aulas.

Aos amigos do Ensino Médio, Weyda, Conceição, Cássio, Francisco e Gean, que apesar da distância mantemos uma amizade sincera e ao meu amigo Wanderson, por sempre estar presente e me apoiar.

Aos amigos que fiz na Engenharia Civil: Gabriel, Kaio, Fred, Israel, Gescilam, Jonas, Matheus, Jesus, Renata, Beliza, Isabel, Mateus, Aldaianny, Altanizio, Diego, David, e em especial ao João Lucas, que me ensinou bastante com sua convivência e momentos de riso na volta para casa.

Ao pessoal do GTTEMA e do OPA-TP pela excelente convivência.

Ao Nelson, por estar presente em algumas reuniões e me ajudar a refletir partes deste trabalho e por dedicar um tempo para participar da banca examinadora.

Aos professores do Departamento de Engenharia de Transportes, pelos conhecimentos compartilhados.

Ao professor Manoel, por dedicar um tempo para participar da banca examinadora e contribuir para melhorar este trabalho.

RESUMO

O transporte coletivo é um modo de transporte essencial para alguns segmentos da sociedade, visto que a elas outros modos não estão acessíveis. Para que ele se torne mais eficiente, são necessárias políticas públicas que o priorizem, promovendo o desenvolvimento sustentável e aumento da sua eficiência e eficácia. Uma das medidas apontadas pela política Nacional de Mobilidade Urbana é a implementação de faixas exclusivas, propiciando aumento da velocidade média, diminuição do tempo de viagem, redução do consumo de combustível e redução custo operacionais e poluição. Contudo, este ganho anunciado pelas empresas de transporte nos indicadores geralmente não segue um método rigoroso de avaliação de políticas públicas, os chamados métodos de inferência causal. Entende-se que diversos fatores podem interferir na evolução de cada indicador, causando viés nessas medições *ex-ante* e *ex-post*. Como oportunidade, há disponível uma rica série histórica de dados de GPS, bilhetagem eletrônica e fiscalização eletrônica que podem ser usados na aplicação de um método de inferência causal, visto que estes geralmente demandam uma grande quantidade de dados longitudinais. Assim, esta pesquisa propõe-se a utilizar esses dados, chamados de big data do transporte público com o objetivo de analisar o real impacto da implantação de faixas exclusivas em Fortaleza em indicadores de desempenho operacional. Para elaborar o método, é feita uma revisão da literatura sobre a priorização do transporte público, métodos de inferência causal e *Big Data* do transporte público. O método proposto está dividido em cinco macro etapas. Na primeira, foi proposto uma integração e consolidação das bases de bilhetagem e de GPS do Sistema Integrado de Transporte de Fortaleza (SIT-FOT), de forma a torna-los adequados ao cálculo de indicadores. Na segunda etapa, é proposto um método de definição de grupo de tratamento e do contrafactual para o caso das faixas exclusivas de Fortaleza. Já na terceira etapa, são definidos os indicadores de desempenho e como calculá-los utilizando os dados de GPS do SIT-FOR. Na quarta etapa, é definido o método de diferença nas diferenças utilizado e por fim, é discutido um modelo com controle para o volume e as limitações dos modelos propostos. Como resultados, verificou-se que o impacto das faixas exclusivas nos indicadores foi bem abaixo do anunciado pela Prefeitura de Fortaleza, ainda que estatisticamente significativos. Além disso, são discutidas as limitações encontradas durante a realização desse trabalho e são dadas recomendações para futuros trabalhos.

Palavras-chave: *Big Data*. Transporte Coletivo. Faixas Exclusivas. Inferência Causal. Diferença nas Diferenças.

ABSTRACT

Public Transport is an essential mode of transport for some parts of the society, since other modes are not accessible to them. For it to become more efficient, public policies that prioritize it are necessary, promoting sustainable development and increasing its efficiency and effectiveness. One of the measures promoted by the National Urban Mobility Policy is the implementation of exclusive bus lanes, providing an increase in average speed, a decrease in travel time, a reduction in fuel consumption and a reduction in operational costs and pollution. However, these gains announced by transport agencies in operational indicators generally does not follow a rigorous method of evaluating public policies, the so-called methods of Causal Inference. Several factors can interfere in the evolution of each indicator, causing bias in these *ex-ante* and *ex-post* measurements. As an opportunity, a rich historical series of GPS, electronic ticketing and electronic surveillance data is available which can be used in a difference in differences causal inference approach, as it generally require a large amount of panel data. Thus, this work proposes to use this data, called the Big Data of Public Transportation with the aim of analyzing the real impact of the implementation of exclusive bus lanes in Fortaleza on operational performance indicators. To elaborate this method, we made a literature review on the Prioritization of Public Transport along with methods of Causal Inference and the Big Data of Public Transport. We divide the proposed method into five steps. In the first, we propose an integration and consolidation of the Fortaleza's ticketing and GPS databases in order to make them suitable for the calculation of indicators. In the second step, we propose a method of defining a treatment and counterfactual group for the case of exclusive bands in Fortaleza. In the third step, we define the performance indicators we used and how to calculate them using the GPS data. In the fourth step, we propose a Difference in Differences approach and finally, we propose a DiD model with control for the volume and we discuss limitations of the proposed models. As a result, we found that the impact of the exclusive bus lanes on the indicators was well below that announced by the City Hall of Fortaleza, although statistically significant. In addition, we discuss the limitations encountered during this work and we give recommendations for future works.

Keywords: *Big Data*. Public Transport. Exclusive Bus Lanes. Causal Inference. Difference in Differences.

LISTA DE ILUSTRAÇÕES

Figura 1 - Mapa das faixas exclusivas implantadas em Fortaleza até o fim de 2018.....	15
Figura 2 - Evolução das faixas exclusivas em Fortaleza	16
Figura 3 - Performance e custo de investimento de sistemas de transporte	24
Figura 4 - Estação de BRT em Curitiba	25
Figura 5 - Premissa dos caminhos paralelos.....	30
Figura 6 - Método	37
Figura 7 - Filtro espacial aplicado em cada linha por Braga (2019)	40
Figura 8 - Filtro Espacial Simplificado	41
Figura 9 - Método de consolidação de Braga (2019)	41
Figura 10 - Fluxograma do método adicional de integração das bases	43
Figura 11 - Evolução das faixas exclusivas com indicação do período escolhido	44
Figura 12 - Corredores de Tratamento e Controle.....	47
Figura 13 - Localização dos equipamentos de fiscalização eletrônica.....	49
Figura 14 - <i>Buffers</i> de controle e de tratamento	51
Figura 15 - Registros de GPS dentro dos <i>Buffers</i>	51
Figura 16 - Tendência das velocidades médias operacionais dos dois grupos.....	54
Figura 17 - Volume mensal nos corredores em maio de 2014	60
Figura 18 - Volume mensal nos corredores do grupo de controle em maio de 2014.....	60
Figura 19 - Volume mensal nos corredores do grupo de tratamento em maio de 2014.....	61
Figura 20 - Volume mensal nos corredores do grupo de controle em março de 2015.....	61
Figura 21 - Volume mensal nos corredores em março de 2015	62
Figura 22 - Volume mensal nos corredores do grupo de tratamento em março de 2015.....	62
Figura 23 - Volume nos corredores entre 6:45 e 7:15 para o mês de março de 2015	63
Figura 24 - Histograma para a velocidade média do grupo de controle em maio de 2014.....	70
Figura 25 - Histograma para a velocidade média do grupo de tratamento em maio de 2014..	71
Figura 26 - Histograma para a velocidade média do grupo de controle em março de 2015	71
Figura 27 - Histograma para a velocidade média do grupo de tratamento em março de 2015	72
Figura 28 - <i>Box plot</i> das velocidades dos dois grupos para maio de 2014 e março de 2015 ...	73
Figura 29 - Teste t para a dupla diferença de velocidade média entre o grupo tratado e o grupo de controle	74
Figura 30 - Teste F para variância das velocidades nos dois grupos.....	75
Figura 31 - Resultado do Modelo com Controle para Volume no Corredor.....	76

LISTA DE TABELAS

Tabela 1 - Amostra de dados de GPS	38
Tabela 2 - Amostra de dados de Bilhetagem Eletrônica	39
Tabela 3 - Corredores de tratamento	45
Tabela 4 - Corredores de controle	46
Tabela 5 - Teste de hipótese dos impactos nos indicadores	56
Tabela 6 - Amostra de indicadores de velocidade média calculados para o grupo de controle em maio de 2014.....	66
Tabela 7 - Amostra de indicadores de velocidade média calculados para o grupo de tratamento em maio de 2014.....	67
Tabela 8 - Amostra de indicadores de velocidade média calculados para o grupo de controle em março de 2015.....	68
Tabela 9 - Amostra de indicadores de velocidade média calculados para o grupo de tratamento em março de 2015.....	69
Tabela 10 - Estatísticas de tendência central e dispersão para os grupos de controle e tratamento em maio de 2014 e março de 2015.	73

LISTA DE ABREVIATURAS E SIGLAS

ADCS	Sistemas de Coleta Automática de Dados
AFC	Coleta Automática de Tarifa
APC	Contagem Automática de Passageiros
API	Interface de Programação de Aplicações
AVL	Localização Automática de Veículos
BRT	Trânsito Rápido de Ônibus
DiD	Diferença nas Diferenças
ETUFOR	Empresa de Transporte Urbano de Fortaleza
GPS	Sistema de Posicionamento Global
GTFS	Especificação Geral de <i>Feed</i> de Trânsito
IPEA	Instituto de Pesquisa Econômica Aplicada
ITSA	Análise de Séries Temporais Interrompidas
PAITT	Plano de Ações Imediatas de Transporte e Trânsito de Fortaleza
PMF	Prefeitura Municipal de Fortaleza
PNMU	Política Nacional de Mobilidade Urbana
PSM	<i>Propensity Score Matching</i>
RD	Regressão com Descontinuidade
SCSP	Secretaria Municipal de Conservação e Serviços Públicos
SIT-FOR	Sistema Integrado de Transporte de Fortaleza
VLT	Veículo Leve sobre Trilhos

SUMÁRIO

1	INTRODUÇÃO.....	12
2	PROBLEMA E PERGUNTAS DE PESQUISA.....	14
3	OBJETIVOS.....	21
3.1	Objetivo geral.....	21
3.2	Objetivos específicos.....	21
4	REVISÃO DA LITERATURA.....	22
4.1	Priorização do Transporte Público.....	22
4.2	Inferência causal.....	27
4.3	<i>Big Data</i> de Transportes e seu potencial.....	32
5	MÉTODO.....	36
5.1	As bases de dados.....	37
5.2	Grupo de tratamento e grupo de controle.....	43
5.3	Indicadores.....	49
5.4	O modelo de estimação do impacto das faixas exclusivas.....	53
5.4.1	<i>Estimativa do impacto na Velocidade Média Operacional.....</i>	54
5.4.2	<i>Estimativa do impacto na variância da Velocidade Média Operacional.....</i>	55
5.4.3	<i>Estimativa do impacto na variância do headway.....</i>	55
5.4.4	<i>Teste de hipótese dos indicadores.....</i>	55
5.5	Controle de variáveis, modelo placebo e considerações finais.....	56
6	IMPACTO DAS FAIXAS EXCLUSIVAS NOS INDICADORES DE DESEMPENHO DO SIT-FOR.....	59
6.1	Determinação do período diário de análise.....	59
6.2	Cálculo dos indicadores.....	63
6.3	Análise Descritiva da Velocidade Média nos Grupos.....	70
6.4	Impacto das Faixas Exclusivas nos Indicadores.....	74
6.4.1	<i>Impacto na Velocidade Média.....</i>	74
6.4.2	<i>Impacto na Variação da Velocidade Média.....</i>	74
6.5	Impacto das faixas exclusivas utilizando DiD com controle para o volume.....	75
7	CONCLUSÕES E RECOMENDAÇÕES.....	77
	REFERÊNCIAS.....	79

1 INTRODUÇÃO

O rápido crescimento populacional aliado à tendência de aumento da população urbana no Brasil dificulta o planejamento urbano adequado, provocando o espalhamento das cidades, a proliferação de zonas com infraestrutura precária e o distanciamento entre moradia e emprego (VINSENTIN *et al.*, 2019). Faltam orientações claras e objetivas das políticas públicas urbanas que promovam a diversificação do uso do solo e acesso a moradias próximas dos centros urbanos para populações carentes. Pelo contrário, há em prática políticas que incentivam o afastamento das populações carentes, como os programas de moradia popular, ao buscarem terrenos mais baratos, geralmente periféricos (NISIDA *et al.*, 2014).

O distanciamento promove a redução das condições de mobilidade e acessibilidade das populações residentes nesses bairros periféricos, causando segregação espacial (IPEA, 2003). São necessários grandes deslocamentos para acessar emprego e lazer, aumentando a necessidade por transporte público e serviços públicos. O transporte tem por função principal integrar as áreas urbanas do ponto de vista espacial, econômico, social e recreativo e isto está intimamente condicionado ao uso do solo das respectivas áreas (BARAT *et al.*, 1973). No entanto, há uma disparidade entre o transporte público e o transporte individual privado.

O transporte tem por função principal integrar as áreas urbanas do ponto de vista espacial, econômico, social e recreativo e isto está intimamente condicionado ao uso do solo das respectivas áreas (BARAT *et al.*, 1973). Assim, o transporte público busca diminuir o abismo entre as classes, ao permitir que as classes menos favorecidas possam realizar seus deslocamentos desejados. No entanto, há uma disparidade entre o transporte público e o transporte individual privado.

Pensando nisso, em 2012 a prefeitura de Fortaleza passou a implantar faixas de circulação exclusiva ao transporte público (FORTALEZA, 2018), visando diminuir os tempos de viagem de seus usuários e, por conseguinte, diminuir a disparidade entre os dois modais. Contudo, é interessante verificar o real impacto dessas intervenções e se produziram o feito esperado, através de indicadores de desempenho do transporte público, visto que a administração eficiente dos recursos públicos depende da correta avaliação da respectiva política (SALAS, 2010).

A Política Nacional de Mobilidade Urbana (PNMU) prioriza os modos de transporte não motorizado e transporte público, sendo um dos princípios a eficiência, eficácia e efetividade na prestação de serviços de transporte urbano, bem como a equidade no acesso ao transporte público, reduzindo as desigualdades e promovendo inclusão social (IPEA, 2013).

Este trabalho está estruturado em uma introdução, onde além do afirmado acima, contém uma contextualização da implantação de faixas exclusivas na cidade de Fortaleza, da avaliação do desempenho de intervenções e do princípio fundamental da inferência causal, a filosofia dos resultados potenciais (*potential outcomes*), a partir das quais propôs-se desenvolver o problema principal de pesquisa e as questões motivadoras.

No capítulo 3 são mostrados o objetivo geral e os objetivos específicos deste trabalho, visando responder o problema de pesquisa e às questões motivadoras.

No capítulo 4 é feita uma revisão da literatura com o objetivo de confirmar as lacunas propostas nos problemas de pesquisa, bem como justificar as questões e afirmações levantadas nesta introdução. Introdutoriamente, é feito uma revisão sobre priorização do transporte público, onde é feito um histórico da implantação de faixas exclusivas, começando pelas primeiras faixas exclusivas implantadas no mundo e chegando até as primeiras faixas exclusivas implantadas no Brasil, bem como seus respectivos objetivos, resultados esperados e resultados observados. Além das faixas exclusivas, outras estratégias de priorização do transporte público são discutidas, brevemente. Em seguida, foi levantado informações sobre o ramo da inferência causal que tem por núcleo o princípio dos resultados potenciais (*potential outcomes*) e os diversos métodos que surgiram desse princípio. É dado ênfase no método da diferença nas diferenças (*DiD*), o qual é visto com um dos mais robustos, dadas às condições necessárias à sua utilização. Por fim, foi feita uma revisão sobre os sistemas automatizados de coleta de dados de transporte (*ADCS – Automated data collection systems*) e a utilização de seus dados para estimativa de indicadores de desempenho. Nessa seção, são explanados os principais ADCSs, seus potenciais em análises de transporte e suas limitações, bem como trabalhos que usaram ADCSs para calcular indicadores de transporte público, com ênfase em trabalhos que os utilizaram para o cálculo de indicadores de desempenho operacional.

Na seção 5 é discutido todo o método utilizado neste trabalho, iniciando pela descrição das bases de dados utilizada e pelos feitos de Braga (2019) em sua dissertação, cujo método utilizado para consolidar os dados do SIT-FOR serviu de base para o método de consolidação empregado neste trabalho, com algumas modificações. Em seguida, é exposto o método utilizado na definição do grupo de tratamento e do grupo de controle, necessário à análise causal. Em seguida é discutido os indicadores utilizados na análise, seguindo pelo modelo de inferência causal utilizado neste trabalho. Por fim, são discutidas as limitações do modelo de inferência causal proposto, bem como é discutido um modelo placebo.

Por fim, na seção 6 são apresentados e discutidos os resultados obtidos, e na seção 7 são apresentadas conclusões e considerações finais sobre o todo deste trabalho.

2 PROBLEMA E PERGUNTAS DE PESQUISA

Para que o transporte público se torne mais eficiente são necessárias políticas públicas que promovam o desenvolvimento sustentável e aumentem a sua eficiência e a sua eficácia. Nesse sentido, uma das medidas apontadas pelo PNMU é a definição de faixas exclusivas para o transporte público e para o transporte não motorizado, propiciando uma divisão mais justa do espaço físico das vias, uma vez que um ônibus transporta em média a mesma quantidade de passageiros que 50 carros (IPEA, 2013). Ainda segundo o PNMU, as faixas exclusivas aumentam a eficiência do transporte público, propiciando aumento da velocidade média, diminuição do tempo de viagem, redução no consumo de combustível, custos operacionais e poluição, podendo até possibilitar redução nas tarifas (IPEA, 2013).

Nesse contexto, há pesquisas que indicam possíveis resultados positivos oriundos da priorização do transporte público, como estudos feitos no Reino Unido, atribuindo um crescimento de 30% na demanda à implantação de faixas exclusivas (UKPTA, 2010), estudos feitos em Nova York, onde se observou redução do tempo de viagem de até 27% e aumento na demanda de 48% (ALRT, 2020). No Brasil, estudos feitos após a implantação do BRT de Curitiba indicaram migração de cerca de 25% dos usuários do transporte individual privado para o transporte público após a implementação do referido sistema de BRT (Rabinovitch e Hoehn, 1995).

A prefeitura de Fortaleza parece ter seguido essa diretriz, visto que, em 2012 a passou a implantar faixas de circulação exclusiva ao transporte público (FORTALEZA, 2018), porém, somente em junho de 2014, com o projeto das faixas exclusivas desenvolvido pela Secretaria Municipal de Conservação e Serviços Públicos (SCSP), por meio do Plano de Ações Imediatas de Transporte e Trânsito de Fortaleza (PAITT) (FORTALEZA, 2018), que a cidade usaria a implantação de faixas exclusivas como uma solução sistemática de priorização do transporte público.

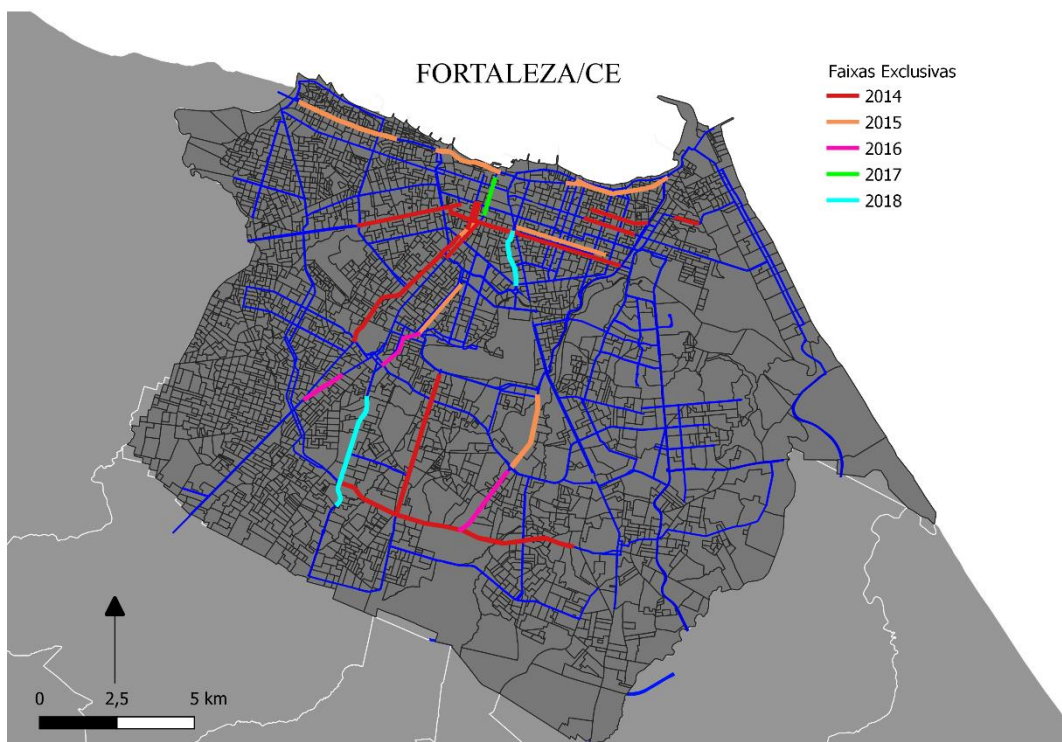
Em 2012 havia duas avenidas em Fortaleza as quais possuíam prioridade ao transporte público: a Avenida Francisco Sá e a Avenida João Pessoa (PORTAL O POVO, 2012). Entretanto, tais prioridades eram do tipo contra-fluxo, ou seja, os ônibus trafegavam em sua faixa exclusiva no sentido contrário ao restante do tráfego. Esse regime de fluxo pode ser perigoso e causar mais acidentes, motivo pelo qual os outros contra-fluxos que existia na cidade foram retirados (PORTAL O POVO, 2012).

Neste mesmo ano, no mês de setembro, foi implantado o primeiro trecho de faixa exclusiva propriamente dito, no mesmo sentido do tráfego, na Rua Padre Ibiapina, no centro da cidade. No entanto, sua extensão se resumia a um trecho de apenas 110 metros.

Fortaleza não veria mais nenhuma implantação de faixas exclusivas até o início de 2014 com a entrega de 6,6km de faixas exclusivas na avenida Bezerra de Menezes, como parte do pacote de obras de mobilidade urbana da cidade para a Copa do Mundo de 2014 (PORTAL MOBILIZE, 2012). No mesmo ano, houve um salto de novos 49,8 km implantados, como parte do referido projeto desenvolvido pela SCSP, por meio do grupo PAITT (FORTALEZA, 2018).

Até dezembro de 2018, Fortaleza já somara 107,4 km de faixas exclusivas, em 33 avenidas da cidade (FORTALEZA, 2018). As faixas exclusivas foram implantadas principalmente em vias arteriais da cidade, com a distribuição espacial demarcada no mapa de vias na figura 1. Além disso, é interessante ver a evolução da quantidade de quilômetros de faixas exclusivas implantadas no gráfico da Figura 2, onde vê-se melhor os saltos nos anos de 2014 e 2015.

Figura 1- Mapa das faixas exclusivas implantadas em Fortaleza até o fim de 2018



Fonte: elaborado pelo autor. Dados da PMF.

Figura 2- Evolução das faixas exclusivas em Fortaleza



Fonte: elaborado pelo autor. Dados: PMF.

Segundo a Prefeitura de Fortaleza (2018), esperava-se que houvesse um crescimento de cerca de 40% na velocidade operacional dos ônibus que circulam nas duas primeiras avenidas beneficiadas pelo projeto do PAITT, em junho de 2014: a Avenida Dom Luís e a Avenida Santos Dumont. O crescimento medido pela prefeitura foi de 143% e 207%, respectivamente. Contudo, ambas as avenidas também foram objeto da implantação de um binário. Por esse motivo, não é possível afirmar que tal crescimento se deu pela implantação das faixas exclusivas. Dessa forma, o desempenho na velocidade operacional medido pela prefeitura representa apenas uma caracterização do cenário *ex-post* intervenção, onde a relação causal entre faixas exclusivas e aumento na velocidade não pode ser afirmada. Isso dito, **verificar o cenário antes e após a intervenção é suficiente para avaliar o efeito das faixas exclusivas?**

O mesmo estudo também fora feito para outros trechos beneficiados pelo projeto, como a Avenida Carapinima, com aumento de 160% e a Avenida da Universidade, com aumento de 70% (FORTALEZA, 2018). Todavia, pelo mesmo motivo exposto anteriormente, essas medidas são uma caracterização do cenário pós intervenção e não se pode atribuir esse aumento em velocidade às faixas exclusivas em específico, pois diversos outros fatores podem interferir.

Embora esse tipo de medição possa ser útil para caracterizar a situação atual de um equipamento do sistema de transporte e possivelmente compará-la com uma situação passada,

a relação causal de uma intervenção no sistema fica comprometida. Nesse contexto, a avaliação de políticas públicas é um ponto central na administração eficiente de recursos públicos, uma vez que, dado a limitação dos recursos públicos, sua aplicação eficiente depende dos resultados causados pela política pública em relação às suas metas (SALAS, 2010). Dessa forma, surge a questão: **quais os benefícios sentidos pelos usuários devido às faixas exclusivas?**

Segundo Schiff *et al.* (2017), os reais efeitos de uma intervenção somente podem ser estimados se os indicadores forem confrontados à uma situação contrafactual: o cenário hipotético onde a intervenção não foi implantada e todos os outros fatores permanecem os mesmos. Contudo, esse cenário é impossível de ser observado, já que apenas o cenário com intervenção ou o cenário sem intervenção se concretizam, não ambos. Dessa forma, o cenário contrafactual deve ser estimado usando dados e modelos estatísticos (Schiff *et al.*, 2017).

A escolha do contrafactual é de suma importância para corroborar o resultado da análise causal. Geralmente, o grupo de controle, ou seja, a estimativa do contrafactual, é escolhido com base nas características do grupo de tratamento, buscando sempre selecionar unidades para o grupo de controle com características similares ao grupo de tratamento. Nesse contexto, fica evidente que é importante saber o critério utilizado pelo poder público para escolha das unidades a receberem o tratamento. Isso nos leva a seguinte pergunta: **quais variáveis ou indicadores podem ter sido usadas na escolha das vias a receberem faixa exclusiva? É possível coletá-las através das bases de dados disponíveis?**

Além da escolha do contrafactual com base em características similares entre os grupos, uma outra situação pode acontecer quando se tem dados observacionais: o chamado quase experimento, que ocorre quando a implantação das intervenções ocorre de forma defasada, desde que todas unidades tratadas tenham sido escolhidas pelo mesmo critério (Campbell e Brakewood, 2017). O quase experimento é preferível à uma análise puramente de dados observacionais, visto que uma relação direta entre o contrafactual e o grupo de tratamento já é estabelecida. Quando um quase experimento não é possível, há técnicas para construir um contrafactual, como a *Propensity Score Matching* (PSM) e a construção de contrafactuais sintéticos (Schiff *et al.*, 2017). Isso posto, surge as seguintes perguntas: **como estimar o contrafactual das faixas exclusivas de Fortaleza? É possível tratar o fenômeno como um quase experimento?**

Outras intervenções foram feitas na rede de circulação viária de Fortaleza, como os binários e o BRT da avenida Bezerra de Menezes, concomitantemente com as faixas exclusivas, em alguns corredores. Sabendo que essas outras intervenções podem afetar os resultados de

análises *ex-post*, pergunta-se: **vias que receberam outras intervenções, além das faixas exclusivas, podem ser incluídas nas análises? Os métodos permitem isso?**

Vários métodos de inferência causal foram propostos na literatura, desde métodos para uma intervenção isolada a métodos para um grupo de intervenções em uma determinada área. Além da intervenção, a aplicabilidade desses métodos é fortemente dependente da quantidade e do tipo de dados disponíveis anteriormente e posteriormente à intervenção, sendo os métodos mais robustos conhecidos pelo grande consumo de dados necessários para serem empregados (Schiff *et al.*, 2017). Dessa forma, **com os dados disponíveis, qual o método ou conjunto de métodos a serem empregados na inferência dos efeitos das referidas implantações? E ao aplicar um método de inferência causal, houve realmente mudanças significativas nos indicadores? Os dados disponíveis permitem atribuir essas mudanças às faixas exclusivas?**

Schiff *et al.* (2017) sustentam que, idealmente, deve-se planejar uma coleta de dados robusta em um ponto no tempo anterior à intervenção e em um ponto no tempo posterior à intervenção para possibilitar a utilização de uma das técnicas mais robustas de inferência causal, bem como de características observáveis que possam ter influenciado na escolha das partes do sistema a receberem as intervenções.

Os métodos de coleta de dados operacionais do transporte público coletivo consolidados e empregados nas últimas décadas são geralmente manuais e necessitam de vários pesquisadores para tornar possível seu emprego (STORCK, 1988). Além disso, necessitam de grande esforço e tempo dos pesquisadores, podendo ser também custosos do ponto de vista econômico e precisam ser cuidadosamente planejados quanto ao momento certo para coleta, uma vez que coletar amostras muito grandes seria impraticável.

Os dados para o cálculo de indicadores de desempenho muito importantes, como a velocidade operacional média e o tempo de viagem, segundo os métodos manuais, precisam de vários pesquisadores posicionados nos pontos de início e término das linhas a anotar uma identificação do veículo e os horários de partida e chegada (STORCK, 1988). Além da dificuldade na coleta, surge também uma dificuldade na tabulação e análise desses dados, havendo muitas brechas para erros humanos, ao copiá-los, por exemplo. Em análises dos efeitos de intervenções no sistema de transporte, geralmente é necessária grande quantidade de dados, em diferentes pontos no tempo, o que pode tornar o emprego de métodos manuais de coleta muito onerosos, quando não impossíveis, uma vez que a coleta de dados anteriores à intervenção nem sempre é planejada pelos agentes públicos.

Com o advento do monitoramento de frotas de veículos via *Global Positioning System* (GPS) e da bilhetagem através de *Smartcard* nos sistemas de transporte público urbano, gera-se continuamente volumes massivos de dados acerca da operação e dos usuários do sistema. Ao contrário dos métodos de coleta manuais, são coletados automaticamente e em tempo real dados quase populacionais e altamente desagregados, de grande potencial para análises de desempenho. Assim, é necessário responder à pergunta: **quais indicadores de desempenho do transporte público podem ser calculados com as bases de dados disponíveis?**

Além disso, pela natureza continuada e automática, há disponível uma rica série temporal de dados, tornando o emprego de técnicas de inferência causal robustas possível, no caso de intervenções no transporte público. Essa oportunidade traz à tona as questões: **quais os indicadores mais adaptados ao uso dessas técnicas de inferência? Qual o nível de agregação dos dados que transparece melhor o efeito das intervenções? Por trecho de via? Por linha?**

A frota de ônibus do transporte público de Fortaleza vem sendo monitorada via GPS desde julho de 2008 por iniciativa do Sindicato das Empresas de Ônibus do Ceará (SINDIÔNIBUS) e da Empresa de Transporte Urbano de Fortaleza (ETUFOR) (SOUSA, 2014). Já a bilhetagem eletrônica, foi implantada em junho de 2014 com o intuito de permitir a integração temporal no sistema (PORTAL DIÁRIO DO TRANSPORTE, 2013). Ambas as bases são disponibilizadas pela prefeitura da cidade desde 2014. Contudo, a bilhetagem eletrônica produz informações agregadas por validação, e não por linha ou veículo. Além disso, há a disponibilidade de dados padronizados da oferta de transporte público, como o GTFS, utilizado notadamente pelo *Google Maps*.

As possibilidades de análises aumentam bastante ao unir essas bases de dados, uma vez que poder-se-ia relacionar diversos atributos, como usuário, linha e sentido, à localização do veículo. No entanto, no caso de Fortaleza, devido a natureza de geração desses dados e ao gerenciamento do sistema, essas bases não são geradas pelo mesmo sistema e apresentam problemas quanto à sua união. Nesse sentido, Braga (2019) realizou um esforço de tratamento e consolidação dessas bases de dados.

Com essas bases de dados consolidada, têm-se dados que podem ser usados para calcular diversos indicadores operacionais do sistema de transporte público de Fortaleza e adentrar em um nível de agregação capaz de fornecer esses indicadores para trechos de vias onde houve implantação de faixa exclusiva, tornando possível empregar uma técnica robusta

de inferência causal para determinar o real impacto das faixas exclusivas nos indicadores do transporte público.

Dessa forma, surge a questão mais abrangente desse trabalho: **qual o impacto das implantações de faixas exclusivas em Fortaleza para os usuários do transporte público?**

3 OBJETIVOS

Nesta seção é apresentado o objetivo geral da pesquisa e os objetivos específicos.

3.1 Objetivo geral

O objetivo geral deste trabalho é avaliar o impacto em indicadores operacionais do transporte coletivo devido à implantação de faixas exclusivas na cidade de Fortaleza, utilizando-se de banco de dados de GPS e bilhetagem do transporte público da cidade, ferramentas de análise de *big data*, oferecendo subsídios para o poder público a fim de auxiliar os gestores na tomada de decisão.

3.2 Objetivos específicos

Os objetivos específicos, alinhados às questões de pesquisa, são:

- Levantar indicadores operacionais de transporte público na literatura para vias sem faixa exclusiva e com faixa exclusiva.
- Determinar os métodos aplicáveis aos dados disponíveis para determinar os efeitos da implantação das faixas exclusivas.
- Estimar um contrafactual para a implantação das faixas exclusivas.
- Aplicar um método ou conjunto de métodos para determinar o real impacto das implantações.
- Propor um método de integração dos bancos de dados de GPS e de bilhetagem do transporte público de Fortaleza de forma consolidar os dados e diminuir os problemas enfrentados por pesquisadores que já usaram essas bases.

4 REVISÃO DA LITERATURA

Neste capítulo é apresentada uma revisão da literatura sobre a priorização do transporte público e como avaliar os impactos dessas intervenções. Na seção 4.1 é apresentada uma breve discussão e histórico da priorização do transporte público no mundo com ênfase nas faixas exclusivas. Na seção 4.2 são apresentados métodos de inferência causal, que são utilizados para estimar o impacto de intervenções feitas em um dado grupo. Nessa mesma seção é dado um pouco mais de enfoque no método da diferença nas diferenças. Passo seguinte, na seção 4.3 é discutido sobre o rastreamento de veículos dos sistemas de transporte público via GPS e demais bancos de dados gerados por sistemas de transporte. Por fim, na seção 4.4 é comentado sobre trabalhos feitos com o intuito de mensurar os impactos de intervenções em sistemas de transporte utilizando-se de bancos de dados de transporte, como dados de rastreamento e de bilhetagem eletrônica.

4.1 Priorização do Transporte Público

O crescimento das cidades e as deficiências no planejamento urbano trazem consigo um problema de mobilidade urbana: as distâncias cada vez maiores que seus habitantes precisam atravessar para realizar suas atividades e, por conseguinte o aumento no tempo gasto em trânsito. Várias estratégias foram adotadas, desde o uso do transporte individual privado, como nos Estados Unidos durante o século XX, ao estímulo ao transporte público coletivo, como em alguns países da Europa.

Ônibus são a tecnologia de transporte público mais utilizada, é um serviço fácil de modificar e de implementar, sendo assim o meio mais econômico para linhas não muito movimentadas (Vukan, 2007). Contudo, segundo Vukan (2007) essa flexibilidade se torna uma desvantagem para linhas mais carregadas, pois lhes faltam a imagem de permanência, confiabilidade e eficiência em transportar grandes volumes que, por exemplo, trens possuem.

Burinskienė (2014) diz que o transporte público é um dos mais importantes modos de transporte urbano, continuando a ser o principal meio de transporte nas cidades, pois assegura o desenvolvimento sustentável de diversas atividades econômicas, sociais e culturais, além de contribuir na diminuição do fluxo nas ruas e na diminuição da poluição ambiental. Várias cidades contam com os chamados sistemas de transporte público em massa, como o Metrô, VLT (Veículo Leve sobre Trilhos) e BRT (*Bus Rapid Transit*) os quais possuem uma característica importante em comum que assegura sua velocidade, capacidade e confiabilidade:

o uso de uma infraestrutura específica, inacessível ao tráfego geral (Deng *et al*, 2011). Assim, Deng *et al.* (2011) defendem a união dessa velocidade e confiabilidade propiciada pela segregação à flexibilidade e ao baixo custo dos serviços convencionais de ônibus urbano.

A maioria dos ônibus operaram em corredores normais, e estando concorrendo com os veículos particulares, sua velocidade e confiabilidade depende das condições do tráfego (Vukan, 2007). Nesse mesmo sentido, Burinskienė (2014) aponta que o número cada vez maior de veículos particulares circulando nas ruas elevam o fluxo a próximo da capacidade, diminuindo a velocidade operacional dos ônibus, um dos principais fatores que desencorajem as pessoas a utilizarem o transporte coletivo.

A velocidade dos ônibus é sempre inferior aos carros por precisarem parar para embarcar e desembarcar passageiros, então uma viagem de ônibus se tornam muito pouco competitiva em relação a uma viagem de carro no mesmo corredor, dado que o tráfego é misto (Vukan, 2007).

Vukan (2007) sustenta que por esse motivo medidas de priorização para os ônibus podem ser adotadas para torná-los mais eficientes e atrativos aos usuários não cativos. Ainda segundo Vukan (2007), as principais priorizações para ônibus são:

- Preferência em interseções semaforizadas: uma faixa separada é disponibilizada para os ônibus nas interseções e recebem o sinal verde para que possam seguir à frente do tráfego.
- Alternação das paradas: paradas logo após ou logo antes de interseções semaforizadas, para que os ônibus que estão no sinal verde possam cruzar a interseção antes de parar.
- Faixas exclusivas: são faixas demarcadas para uso exclusivo do transporte coletivo.
- BRTs: são estruturas segregadas fisicamente do corredor somente para ônibus. Requer grande investimento e, por isso, é mais usado apenas em seções de linhas ônibus com volume muito grande, concorrendo diretamente com trens.
- Faixas reservada a veículos com alta ocupação: são faixas reservadas para carros com mais de um passageiro e ônibus. Similar as faixas exclusivas para ônibus, mas não passam uma imagem aos passageiros de uma priorização como as faixas exclusivas.

Oliveira Neto (2009) investigou a performance do uso de técnicas passivas e ativas de priorização de semáforos para ônibus (ciclos fixos e controlados em tempo real), em uma via arterial de Fortaleza, a avenida 13 de Maio. As principais conclusões foram que os

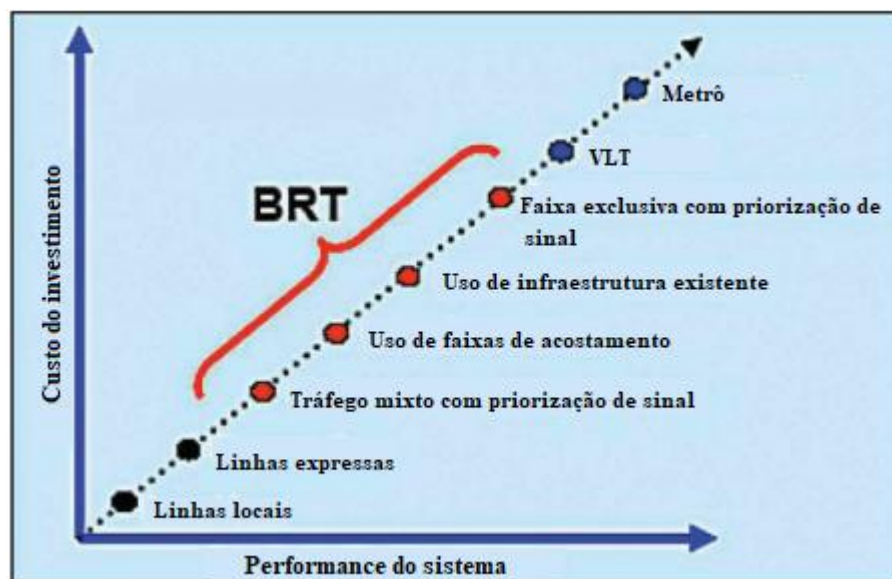
resultados não favoreceram a implementação de técnicas passivas de priorização, para isso, Oliveira Neto (2009) aponta como possibilidade a existência de tráfego misto na via, e que a técnica mais adequada de priorização de sinais foi a em tempo real, desde que precisamente configurada para priorizar o tráfego na via principal e acomodar a aleatoriedade do tráfego nas vias transversais.

A escolha do número de paradas de ônibus e sua localização é uma tarefa complicada que depende da probabilidade de o ônibus parar, velocidade, *headway* e saturação das vias (Tirachini, 2014). Tirachini (2014) conduziu um estudo prático em Sydney e seu principal achado é de que a velocidade dos ônibus aumenta consistentemente com o aumento do distanciamento entre paradas, mesmo em cenários de alta demanda, como BRTs.

Deng *et al.* (2011) afirmam que a implementação de BRTs emergiu no início do século XXI como o modo de transporte urbano de passageiros líder nas principais cidades de todos os continentes, devido a habilidade de implementar transporte de passageiros em massa rapidamente, à custo moderado e aproveitando recursos existentes, bem como influenciado por instituições de financiamento, como o Banco Mundial.

Na figura 3 é apresentado um esquema mostrando a performance do sistema de transporte em relação ao seu custo de investimento.

Figura 3 - Performance e custo de investimento de sistemas de transporte



Fonte: Adaptado de Deng *et al.*, 2009.

Segundo Deng *et al.* (2011), o conceito moderno de BRT foi desenvolvido nos anos 1970 por especialistas em planejamento de transporte latino-americanos, explorando a rápida e não tão cara forma de alavancar a velocidade do ônibus.

Curitiba é frequentemente citada como exemplo bem-sucedido de implementação de BRTs, que operam quase como o transporte metroviário de superfície, utilizando faixas dedicadas, infraestrutura de estações, alta velocidade e confiabilidade, mas a um custo muito menor (Smith e Raemaekers, 1988). Uma pesquisa conduzida em 1991 sugeriu que cerca de 25% dos proprietários de automóveis passaram a também utilizar o sistema (Rabinovitch e Hoehn, 1995). Um exemplo de infraestrutura de BRT implantado em Curitiba é mostrado na figura 4.

Figura 4- Estação de BRT em Curitiba



Fonte: PORTAL ARCHDAILY (2020)

Como Deng e Nelson (2011), Vukan (2007) também aponta que os melhores sistemas de BRT são encontrados em Curitiba e Ottawa. Contudo, Vukan (2007) pondera que a implementação de BRTs precisa de um investimento razoável e sustenta que as faixas exclusivas são a medida de priorização mais eficaz, visto que segrega os ônibus do tráfego.

Esse pensamento na comunidade científica não é recente, pois a primeira faixa exclusiva de ônibus implantada em uma cidade remonta à década de 30, na cidade de Chicago, em 1939 (Deng *et al.*, 2011).

A primeira faixa exclusiva ao transporte público no continente europeu fora implantada em 1963, em Hamburgo, na Alemanha, cujo sucesso estimulou a adoção dessa estratégia em outras cidades europeias (ALRT, 2020). Já no ano seguinte, foram implantadas em Paris e em 1968, Londres.

De início, a adoção das faixas exclusivas não fora bem aceita pelos usuários de carros na maioria dos países europeus (Burinskienė, 2014). Contudo, segundo Burinskienė (2014), um estudo conduzido em Turin, na Itália mostrou que 51% dos proprietários de carros particulares estariam dispostos a usar o transporte público, caso faixas exclusivas fossem implantadas na cidade. Outro estudo da época, feito em Birmingham, Reino Unido, mostrou que houve um aumento de 30% nos usuários do transporte público da cidade, após a implantação das faixas exclusivas (UK Public Transport Authority, 2010). Assim, essas pesquisas sugerem que a qualidade e eficiência no transporte público em muito contribui para a migração dos usuários do transporte individual privado para ele, gerando benefícios sociais, econômicos e ambientais.

Devido ao sucesso da França e do Reino Unido nesse aspecto, governos de outros países e cidades passaram a também privilegiar as faixas exclusivas, e em 1974, mais de 100 cidades já contavam com faixas exclusivas em operação (ALRT, 2020).

No continente americano, foi feito um experimento na linha de ônibus M15 de Nova York, que operava a uma velocidade média de 8,53 km/h e para melhorar a situação, duas faixas exclusivas de ônibus segregadas foram instaladas, uma na Primeira Avenida e outra na Segunda Avenida. Como resultado, o tempo médio de viagem foi reduzido em cerca de 21% a 27% e quantidade de passageiros transportados aumentou em 48% (ALRT, 2020).

Também contam com faixas exclusivas as cidades de Berlin, na Alemanha, reconhecida pelos ônibus pontuais e, mais recentemente, Moscou, na Rússia, vem implantando faixas exclusivas em escala desde 2010 (ALRT, 2020).

Um consenso nos trabalhos revisados é de que as faixas exclusivas aumentam a velocidade operacional dos ônibus e a sua confiabilidade (Deng *et al.*, 2011; Vukan, 2007). Dessa forma, indicadores que reflitam a velocidade e confiabilidade dos ônibus, como a velocidade média operacional e o *headway* são escolhas naturais de indicadores em estudos sobre o tema.

4.2 Inferência causal

Os métodos tradicionais de inferência causal vêm sendo usados em pesquisas no campo de transportes, desde investigação do efeito causal de redução de velocidades de tráfego em segurança, impactos na demanda por ônibus devido a sistemas de bicicletas compartilhadas e implantação de faixas exclusivas de ônibus (SCHIFF *et al.*, 2017; CAMPBELL, 2017; PERCOCO, 2014).

Segundo Schiff *et al.* (2017), os seguintes problemas devem ser considerados ao tentar estimar os efeitos causados por uma intervenção:

- O tratamento deve ser claramente identificado bem como deve ser determinado precisamente em que ponto no tempo o tratamento foi aplicado e os efeitos esperados;
- Definição das unidades observacionais, ou grupos que receberam o tratamento;
- Definição de indicadores que se espera serem afetados pela intervenção;
- Definição do contrafactual, ou no caso de experimentos não randomizados, a estimativa desse, o grupo de controle.
- O método para comparar a diferença de resultados entre os dois grupos;
- Deve haver uma quantidade suficiente de observações para distinguir o efeito da intervenção de variações aleatórias nos dados;
- A probabilidade de uma unidade ser escolhida para receber a intervenção ser afetada por variáveis observáveis e não observáveis;
- As características dos grupos de controle e de tratamento devem ser suficientemente similares;

Para se estimar os efeitos causados por uma intervenção é preciso que haja um contrafactual, ou seja, o estado do mundo sem a intervenção. Porém, o contrafactual é impossível de ser observado diretamente, então ele deve ser estimado usando dados e ferramentas estatísticas (Schiff *et al.*, 2017).

Ainda segundo Schiff *et al.* (2017), o problema com análise do tipo antes e depois da intervenção é que ela não permite retirar dos resultados outros efeitos decorrentes de outras variáveis não relacionadas com a intervenção bem como torna-se difícil justificar os resultados observados como resultantes da intervenção.

A fundação de todos os métodos de inferência causal é modelo causal de Rubin, ou modelo dos resultados potenciais (Schiff *et al.*, 2017).

Segundo Rubin (1974, p. 689, tradução nossa),

Seja y o efeito de uma intervenção medido em uma unidade observacional i que pode ter sido sujeita a tratamento (y_1) ou não (y_0). Para a unidade i é possível definir o resultado potencial y_i^0 , o cenário em que a unidade i não recebeu o tratamento, e o resultado potencial y_i^1 , o cenário em que a unidade i recebeu o tratamento. O efeito médio do tratamento (ATE) é dado pelo valor esperado do efeito do tratamento em todas as unidades i : $ATE = E[y_i^1 - y_i^0]$ (1)

Esse modelo considera que ambos os cenários, com e sem intervenção na mesma unidade podem ser observados. Porém, isto é impossível, e é chamado de o problema fundamental da inferência causal (Schiff *et al.*, 2017). Para contornar o problema fundamental, diversos métodos foram sugeridos para estimar o efeito do tratamento pelo que pode ser observado.

O primeiro deles é o experimento completamente aleatório, em que as unidades de controle e de tratamento são escolhidas aleatoriamente, permitindo uma comparação de unidades similares dos dois grupos (Schiff *et al.*, 2017). Nesse caso, o grupo de controle escolhido aleatoriamente é a estimação do contrafactual.

Schiff *et al.* (2017) em um relatório para Agência de Transportes da Nova Zelândia sumarizou as principais técnicas de inferência causal, que são as seguintes.

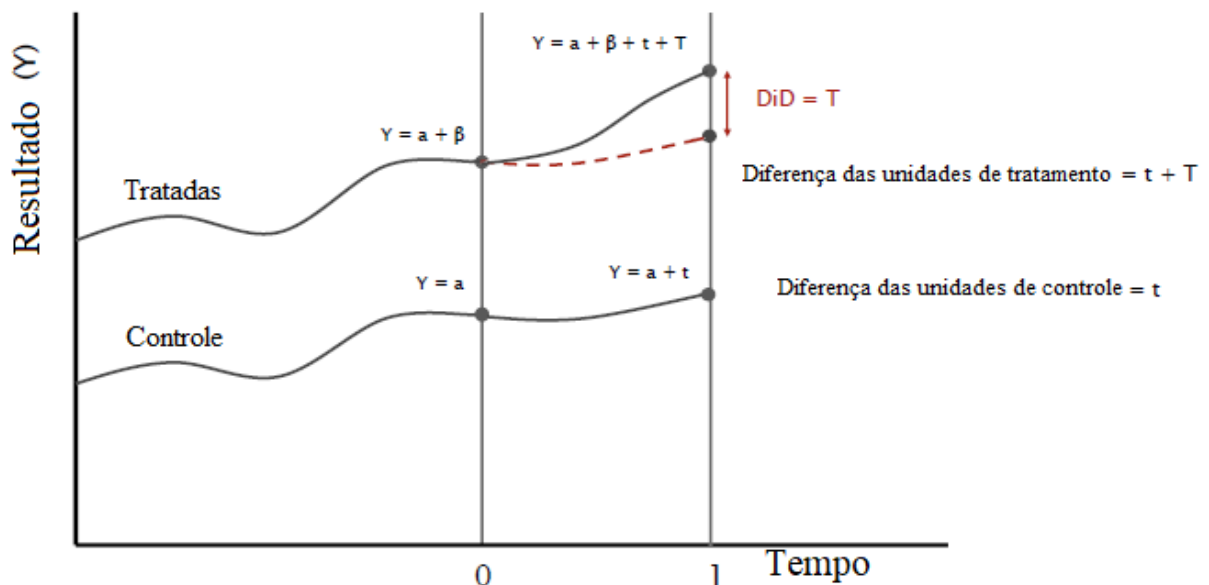
- **Inferência causal com modelos de regressão simples:** os efeitos de uma intervenção podem ser estimados usando um modelo de regressão em dados longitudinais. Contudo, Schiff *et al.* (2017) aponta que essa técnica não produz resultados confiáveis caso os dados não sejam de experimentos aleatórios.
- **Propensity Score Matching (PSM):** PSM é mais usado para definir os grupos de controle e tratamento via um indicador, o PSM, e depois aplicar um modelo de regressão simples. Esse método necessita de grandes quantidades de unidades e pode resultar em um grande número de unidades descartadas da análise (Schiff *et al.*, 2017);
- **Diferença nas diferenças (DiD):** A lógica da DiD vem da premissa de que as unidades de tratamento e de controle seguem caminhos paralelos, e a diferença entre os dois caminhos é devido à intervenção. Essa é considerada uma das técnicas mais robustas de inferência causal (Schiff *et al.*, 2017);

- **Variáveis explanatórias:** são variáveis adicionadas no modelo de regressão para controlar quando uma ou mais variáveis estão correlacionados como erro padrão do modelo de regressão;
- **Regressão com descontinuidade (RD):** é usada quando as unidades de tratamento são claramente escolhidas com base em variáveis observáveis. Contudo, como o modelo é uma extrapolação, são necessárias fortes premissas que tornam o modelo menos confiável (Gelman e Hill, 2006);
- **Análise de séries temporais interrompidas (ITSA):** Os dados das unidades tratadas são observados por regressão com uma interrupção no momento das intervenções, onde aparece uma descontinuidade, atribuída ao efeito da intervenção. Contudo, ITSA requer a extrapolação das tendências além dos períodos com dados observados, o que a torna menos robusta do que outras técnicas (Schiff *et al.*, 2017).
- **Métodos Bayesianos:** são métodos utilizados quando não há grupo de controle suficientemente grande. Contudo, esses métodos possuem dois grandes problemas. Primeiramente, podem ser bastante imprecisos devido ao reduzido número de observações. Segundamente, esses métodos carregam naturalmente um viés na escolha das unidades de tratamento (Schiff *et al.*);

Schiff *et al.*, (2017) sustentam que DiD é o método mais robusto a ser aplicado quando se dispões de dados transversais e longitudinais, ou seja, dados em dois ou mais pontos no tempo simultaneamente para o grupo de tratamento e o grupo de controle. Como esse é o caso neste trabalho, DiD será explicada com mais afinco.

Graficamente, a premissa dos caminhos paralelos pode ser representada conforme a figura 5.

Figura 5- Premissa dos caminhos paralelos



Fonte: Modificado de Gertler *et al* (2010)

As variáveis mostradas no gráfico são:

- a é uma constante comum aos dois grupos;
- β é uma variável que afeta especificamente o grupo de tratamento;
- T é o efeito da intervenção;
- t é o efeito do tempo nos dois grupos;

Formalmente:

$$T = (E[y_1^1 - y_1^0] - E[y_0^1 - y_0^0]) \quad (2)$$

Pela figura 5, vê-se que após calculadas as duas diferenças, o único efeito restante é T , o efeito da intervenção.

A DiD também pode ser aplicada em uma regressão que estima a diferença entre os dois grupos. Isso é especialmente útil quando não se pode observar a variável de interesse diretamente ou se suspeita que alguma variável afeta o grupo de tratamento de uma forma diferente após a intervenção. A formulação da DiD para modelos de regressão é a seguinte:

$$y_{1i} - y_{0i} = \alpha + \beta D_i + \varepsilon_i \quad (3)$$

Onde:

- Y_{1i} é o indicador observado no grupo de tratamento antes da intervenção ($i=0$) e depois da intervenção ($i=1$);

- Y_{0i} é o indicador observado no grupo de controle antes da intervenção ($i=0$) e depois da intervenção ($i=1$);

- D_i é uma variável binária representando se a unidade recebeu tratamento ou não;

- ε_i é o resíduo da regressão;

- α é o intercepto da regressão;

O coeficiente β carrega a diferença devido á intervenção procurada.

A DiD pode ser usada em conjunto com a PSM para diminuir a chance de variáveis não observáveis influenciarem no grupo de tratamento. Os grupos são escolhidos utilizando a PSM e depois é aplicado a DiD. Contudo, essa abordagem requer uma quantidade enorme de unidades e de dados (Schiff *et al.*, 2017).

Uma outra abordagem usada para diminuir essa probabilidade é usada em unidades que receberam o tratamento de forma defasada, pois ambos os grupos carregam as mesmas variáveis implícitas na escolha. Essa foi a abordagem utilizado por Campbell e Brakewood (2017). Nessa abordagem é conhecida como quase-experimento, e é válida desde que as unidades que receberam o tratamento defasado tenham características similares, como é o caso de intervenções que depende de infraestrutura (Campbell e Brakewood, 2017).

4.3 *Big Data* de Transportes e seu potencial

Atualmente os veículos do transporte público, normalmente, contam com sistemas automatizados de coleta de dados de transporte (ADCS – *Automated Data Collection Systems*), notadamente a localização automática de veículos (AVL – *Automated Vehicle Location*), que é coletada por equipamentos de monitoramento GPS, e a Contagem Automática de Passageiros e Cobrança Automática de Tarifa (APC – *Automated Passenger Counting* e AFC – *Automated Fare Collection*) (Pangilinan *et al.*, 2008).

O sistema de posicionamento global (GPS) foi desenvolvido pelo departamento de defesa estadunidense originalmente para aplicações militares, sem perspectivas de torna-lo aberto ao público civil. Entretanto, após um avião de passageiros de uma companhia área sul Coreana ser derrubado pela União Soviética em 1983 após entrar acidentalmente em espaço aéreo soviético restrito, ficou claro que a tecnologia poderia ter salvo aquele voo. No mesmo ano, o presidente estadunidense Ronald Reagan tornou a tecnologia disponível para todo o mundo, embora tenha sido adicionado ruído no sinal, diminuindo a precisão para inferior à 100m, o que impossibilitava diversas aplicações comerciais (*US Dept of Defense*, 2018). Em 2000, um decreto do presidente Bill Clinton liberou o sinal de GPS preciso, possibilitando sua aplicação em diversas áreas da economia, como na construção civil, no transporte, no setor aeroespacial e em dispositivos móveis (*US Dept of Defense*, 2018).

Ainda segundo o departamento de defesa estadunidense (2018), outras agências também têm desenvolvido seus próprios sistemas de posicionamento global, como o russo GLONASS, o chinês BeiDou e o europeu GALILEO.

Nos sistemas de transporte, o GPS e os demais sistemas são utilizados no monitoramento e controle em tempo real de veículos, na localização de emergência e em priorizações de sinais (TOWNES *et al.*, 1997). Essa adoção do GPS nos sistemas de transporte possibilitou a geração de dados massivos de localização dos veículos, permitindo o cálculo de indicadores de transporte como a velocidade operacional, frequência, *headway*, confiabilidade, entre outros (BRINCKERHOFF, 2013).

Uma outra tecnologia que possibilita a geração de dados massivos de transporte é o *smartcard*, tecnologia desenvolvida na metade do século 20 e utilizada no planejamento de transportes desde então (PELLETIER *et al.*, 2011). Ainda segundo PELLETIER *et al.*, o *smartcard* é utilizado mundialmente para coleta de tarifa de transporte público e propiciando uma coleta contínua de dados referentes aos usuários do transporte público. Cada validação de

um usuário deixa diversos registros, como a hora da validação, o tipo de pagamento, a linha e sentido utilizado.

No caso do SIT-FOR, há trabalhos como o de Oliveira (2019), Braga (2019) e Rabay (2017). O sistema de bilhetagem eletrônica (AFC e APC) não é unificado com o sistema de localização automática dos veículos (AVL), fazendo com que as informações provenientes da bilhetagem não possuam localização e as informações provenientes do sistema de localização não tenha informações referentes à linha e sentido. Nesse caso, é necessário a integração dos dados das duas bases, como o feito por Braga (2019), para o SIT-FOR.

Braga (2019), se baseou nos trabalhos Zhao *et al.* (2008), o primeiro autor a explicitar o método de integração das bases de dados, utilizando uma coluna comum às duas bases, *busid*, e buscando incorporar as bases com base nessa coluna e no horário mais próximo. Braga (2019) utilizou uma abordagem similar para georreferenciar os dados de bilhetagem, embora precauções adicionais tenham sido tomadas devido ao fato de no sistema de Fortaleza ser possível um usuário subir no veículo e não validar imediatamente.

Os dados de AVL são, geralmente, difíceis de manusear devido ao grande volume. Esses dados não possuem as informações referentes à linha que o ônibus está servindo, tornando necessário a aplicação de um método de integração com a base de bilhetagem. Braga (2019) propôs um método que foi capaz de integrar cerca de 85% dos registros. Uma grande limitação foi a discrepância entre a identificação do veículo na base de bilhetagem e a identificação do mesmo veículo na base de GPS, que utilizavam códigos diferentes. Para relacionar as duas bases de dados, Braga (2019) utilizou uma base contendo a conversão entre os códigos de cada base, chamada de *dicionário*. Porém, discrepâncias nessa base de dados fizeram com que a integração das bases não fosse completa.

Diversos autores fizeram o esforço de calcular indicadores operacionais através dos dados de AVL, como Quiroga e Bullock (1998) e Cortés *et al.* (2011). No estudo de Queiroga e Bullock (1998), veículos particulares equipados com equipamentos de GPS foram utilizados para a coleta e foi definido um método para consolidação e tratamento desses dados. Devido à grande quantidade de registros (as posições eram medidas à cada 1 segundo), os autores propuseram agregar os dados por link das vias, pontos pré-selecionados de interesse, como cruzamentos, e a utilização de ferramentas de processamento de dados geográficos (GIS) para associar os dados aos pontos criados.

Já no estudo de Cortés *et al.* (2011), os dados de localização da frota de veículos do transporte público de Santiago no Chile foram processados transformando os dados em

tipologia de pontos para a tipologia de *linestrings* (linhas), em seguida alocando as *linestrings* criadas com o itinerário georreferenciado de cada linha do sistema.

Um outro avanço foi trazido nos trabalhos de Arbex e Da Cunha (2016) e Rabay (2017), os quais se utilizaram dos dados de GTFS para integrar as bases e para estimar indicadores operacionais.

Arbex e Da Cunha (2016) tiveram como objetivo fazer uma análise *ex-post* de faixas exclusivas na cidade de São Paulo, utilizando os dados de GPS da frota para 20 dias úteis de agosto de 2013 e 20 dias úteis de abril de 2015. A metodologia utilizada pelos autores foi de integrar os dados de GPS com os dados de GTFS para associar cada registro à uma parada de ônibus, se utilizando de filtros espaciais para isso.

Os dados utilizados por Arbex e Da Cunha (2016) continham, além da localização do veículo, o horário do registro e a linha à qual o veículo estava servindo. Isso possibilitou aos autores seguirem diretamente para o cálculo dos indicadores, uma vez que a integração com uma base de dados de bilhetagem não era necessária. Os autores decidiram avaliar a mediana das velocidades das viagens das linhas da cidade, em toda sua extensão. As linhas de ônibus foram divididas em dois grupos: as que tiveram ao menos um trecho beneficiado com faixas exclusivas e o grupo de linhas que não tiveram nenhum trecho beneficiado com faixas exclusivas.

Para justificar a escolha de grupos de linhas feita por Arbex e Da Cunha (2016), os autores analisaram a participação da demanda dos usuários nos dois grupos. Os autores encontraram que aproximadamente 80% da demanda estava no grupo de linhas beneficiado pelas faixas exclusivas, enquanto as que não o tiveram, apenas 20%. Visto essa discrepância, os autores decidiram selecionar para os grupos linhas de demanda similar, totalizando 240 linhas para o grupo beneficiado pelas faixas exclusivas e 68 para o outro grupo. Com os dois grupos definidos, foi calculado os tempos de viagem, entre a segunda e a penúltima parada, para cada linha através de um algoritmo desenvolvido pelos autores e ao relacioná-lo com a distância do itinerário no GTFS de cada linha, os autores determinaram a velocidade média (Arbex e Da Cunha, 2016).

Arbex e Da Cunha (2016) reconhecem que outros fatores exógenos, como obras nas vias, alterações nos sistemas viários, acidentes e programação semaforica podem ter interferido nos resultados, tornando-o inválido em magnitude, mas válido quando à compreensão do fenômeno e avaliação da tendência geral de evolução das velocidades média dos veículos do transporte público de São Paulo. Os autores constataram que não houve variação estatisticamente significativa da mediana da velocidade média para o grupo que não foi beneficiado com as faixas exclusivas, enquanto o grupo de linhas em que houve implantação

de faixas exclusivas teve variação significativa de 9,13% em média. Em relação ao trabalho de Arbex e Da Cunha (2016), observa-se também que não foi utilizado um método inferência estatística.

As bases de dados de AVL, AFC e APC são frequentemente chamadas de *big data* de transporte pelos autores. O termo *big data* está em alta nos últimos anos, porém, possui definições diferentes na academia, na indústria e na mídia e essa falta de consenso produz ambiguidades e dificulta o uso do termo (Ward, 2013). Esse termo está popularmente associado com apenas o volume de dados, no entanto, este não é o único parâmetro a ser considerado.

Cartner (2001) propôs o uso do termo condicionado a três parâmetros, os três “V”s: volume, velocidade e variedade. No entanto, uma quantidade ou um modo quantitativo de defini-lo não foi proposto. Mais recentemente, um quarto “V” foi adicionado às condicionantes, a veracidade, refletindo o quanto o analista confia ou não nos dados (Cartner, 2012).

Já a *National Science Foundation* (2011) salienta que *big data* é geralmente proveniente de sensores, instrumentos automatizados, e operações na internet, criando bases de dados volumosas, complexas e longitudinais. Nesse contexto, dados provenientes de sensores de GPS em frotas de transporte público e os gerados por cartões de viagem são considerados *big data* (KITCHIN, 2013).

5 MÉTODO

Esta seção apresentará o método utilizado neste trabalho, a fim de atingir seus objetivos. Para isso, foi necessário preparar e adequar os dados disponibilizados pela Prefeitura Municipal de Fortaleza (PMF), relativos ao período compreendido entre 2014 e 2015, de forma a possibilitar os cálculos necessários à obtenção dos indicadores e à aplicação do método de inferência causal das diferenças nas diferenças.

O principal problema de natureza relacional nas medições de GPS da frota reside no fato de que esse banco de dados não possui as informações sobre linha e sentido para cada registro, as quais, em sua ausência, tornariam as análises propostas menos significantes. Para contornar esse problema, Braga (2019) cruzou os registros de GPS com os registros de bilhetagem (*smartcard*). Contudo, essa abordagem expôs uma inconsistência entre as duas bases: cerca de 15% dos registros não puderam ser identificados por esse método, em maior número os referentes às linhas expressas e ao transporte complementar (*topics*).

Nesse contexto, foi proposto uma etapa adicional de consolidação com os dados não identificados remanescentes após aplicação do método de Braga (2019), que não depende da base de bilhetagem. Ressalta-se que este método não foi aplicado à base de dados completa do GPS devido ao grande esforço computacional exigido. O método completo de consolidação dos dados e uma contextualização sobre as bases de dados utilizadas será explicado com maior afinco na seção 5.1.

Após consolidada a base de dados, foi definido os corredores em que foram implantadas faixas exclusivas e corredores que não receberam faixas exclusivas, o grupo de controle, necessário ao método das diferenças nas diferenças. Passo seguinte, definiu-se indicadores que se acreditava serem representativos dos impactos das faixas exclusivas. Para definir o período do dia analisado, utilizou-se dos dados de volume fornecidos por fiscalização eletrônica, também disponibilizados pela PMF. O método de escolha desses corredores será descrito na seção 5.2. Como acredita-se que as faixas exclusivas tenham maior impacto durante a hora de pico, foi proposto também nessa seção um método para identificar a hora de pico do corredor utilizando dados de fiscalização eletrônica.

A escolha dos indicadores é explicitada na seção 5.3, juntamente com o método utilizado para calcular a velocidade média dos veículos no corredor utilizando os dados de GPS.

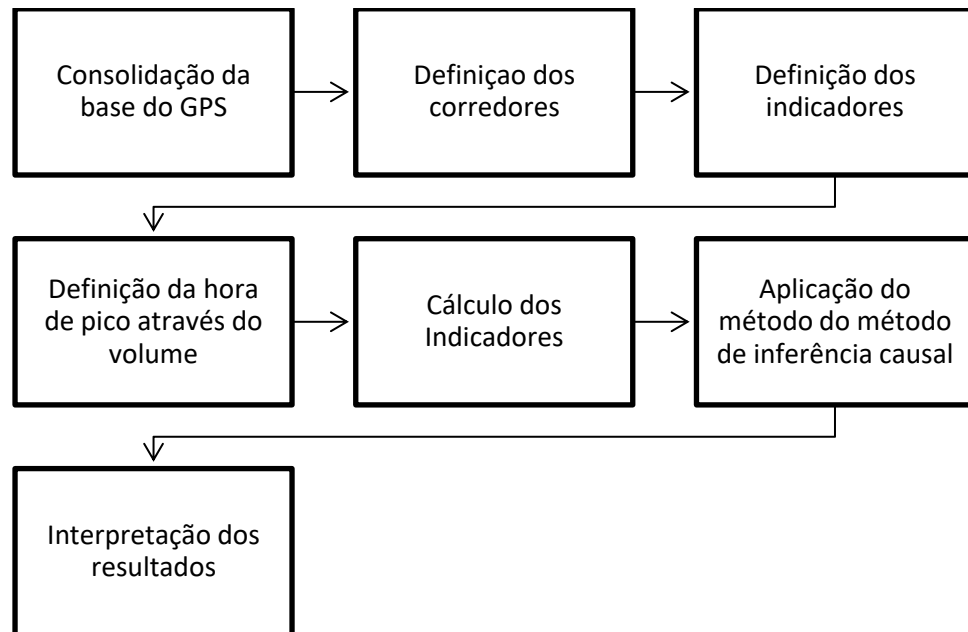
Por fim, na seção 5.4 é mostrado o método das diferenças nas diferenças aplicado nesses indicadores a fim de identificar o impacto da implantação das faixas exclusivas. Além disso, na seção 5.5 é discutido um modelo adicional aplicando o método da diferença nas diferenças utilizando uma regressão linear para a velocidade em função do volume, a fim de

isolar possíveis alterações de volume entre os cenários, que poderiam enviesar a análise, bem como são discutidas algumas limitações do método proposto.

O processamento dos dados foi feito utilizando linguagem *R* e os mapas foram feitos utilizando o *software* QGIS.

A macroseqüência do método está esquematizada na figura 6.

Figura 6- Método



Fonte: elaborado pelo autor.

5.1 As bases de dados

A base principal de dados utilizada neste trabalho foi a base de monitoramento da frota do transporte público de Fortaleza, para o mês de maio de 2014 e março de 2015, disponibilizados pela Empresa de Transporte Urbano de Fortaleza (ETUFOR). Entretanto, como já discutido, é necessário integrar a base de GPS com a base de Bilhetagem Eletrônica para ter dados com as informações das linhas. Para isso, foi utilizado também os dados de Bilhetagem Eletrônica e de GTFS fornecidos pela ETUFOR. Foi também utilizado os dados de Fiscalização Eletrônica para determinar os volumes de veículos nas linhas em estudo.

Os dados brutos de monitoramento GPS da frota de transporte público de Fortaleza foram pré-processados por Braga (2019), os quais possuíam mais de 40 atributos, em sua maior parte nulos ou vazios e foram reduzidos às seguintes colunas:

- *Id*: identificação única à cada registro.
- *Vehicle_vehicleid*: identificação única de cada ônibus.
- *Hora*: dia e horário de cada registro.

- *Latitude*: latitude onde se encontrava o veículo no momento do registro.
- *Longitude*: a longitude onde se encontrava o veículo no momento do registro.
- *Metrictimestamp*: o dia e horário de cada registro em formato UNIX.

As colunas *id* e *metrictimestamp* podem ser suprimidas sem prejuízos à análise proposta nesse trabalho. A coluna *vehicle_vehicleid* pode ter seu nome simplificado para apenas *vehicleid*. Uma amostra dos dados é apresentada na tabela 1.

Tabela 1 - Amostra de dados de GPS

	Hora	Longitude	Latitude	Id Veículo
1	20:59:57	-38,606093	-3,719291	39511
2	20:59:38	-38,502733	-3,803141	32841
3	20:59:33	-38,545281	-3,744476	32068
4	20:59:13	-38,606095	-3,719288	39511
5	20:59:08	-38,502788	-3,803146	32841
6	20:59:03	-38,545281	-3,744476	32068
7	20:59:03	-38,542106	-3,720343	33586
8	20:58:43	-38,6061	-3,7193	39511
9	20:58:38	-38,502748	-3,80311	32841
10	20:58:33	-38,545281	-3,744476	32068

Fonte: O autor.

Como se pôde ver na amostra dos dados mostrada na tabela 1, não há a informação referente a linha e o sentido da viagem em cada registro de GPS. Como é necessário identificar os registros referentes às linhas que passam nos corredores selecionados para análise, propõe-se recuperar essas informações da base de dados de bilhetagem. Para isso, procurou-se uma informação que pudesse ser usada para cruzar os dados das duas bases. Na base de GPS há o campo *Vehicleid* que pode ser usado juntamente com o campo *prefixo_carro* da base de bilhetagem para recuperar as informações da linha e do sentido da viagem. Na tabela 2 é mostrado uma amostra da base de dados de bilhetagem.

Tabela 2 - Amostra de dados de Bilhetagem Eletrônica

	id	hora	linha	prefixo_carro	sentido_viagem	tipo_cartao	integracao
1	0	22/10/2018 16:33	855	30033	Volta	Inteira	N
2	7334274	22/10/2018 16:33	855	30033	Volta	Vale Transporte	S
3	4997848	22/10/2018 16:33	855	30033	Volta	Vale Transporte	N
4	7369510	22/10/2018 16:33	855	30033	Volta	Vale Transporte	N
5	7458128	22/10/2018 16:34	855	30033	Volta	Vale Transporte	N
6	4917679	22/10/2018 16:35	855	30033	Volta	Vale Transporte	N

Fonte: O Autor

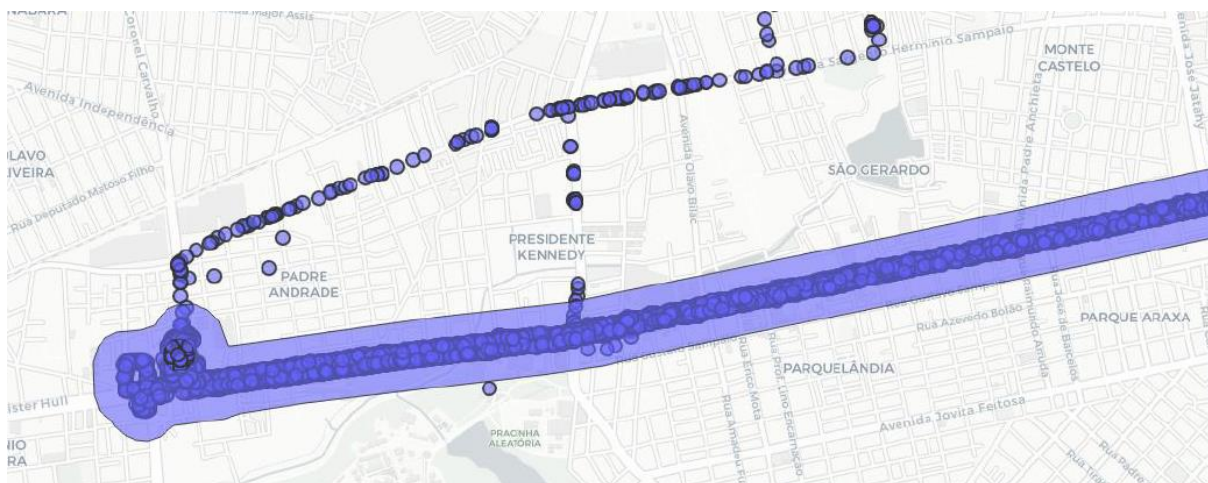
Contudo, o número do *Vehicleid* difere do número do *prefixo_carro*. Essa é a principal causa da limitação na integração das duas bases. Para contornar esse problema, há uma base de dados chamada de *dicionário* que apresenta o respectivo *prefixo_carro* em relação ao *vehicleid* na base de dados de rastreamento da frota. Todavia, o arquivo *dicionário* possui falhas, não possui todas as relações entre ambos os campos, bem como algumas relações contidas possivelmente podem ser válidas para um certo período do ano, e não mais válidas em outros, devido a, por exemplo, veículos que deixam de operar e veículos novos que não são registrados no arquivo ou trocam de equipamento de bilhetagem ou de equipamento de monitoramento GPS, mudando a respectiva numeração. Braga (2019) ao integrar as duas bases para o ano de 2018 observou que aproximadamente 15% dos registros não puderam ser relacionados.

Para aumentar a confiança de que as relações contidas no *dicionário* eram válidas, e principalmente devido a um mesmo veículo poder servir à mais de uma linha no mesmo dia, Braga (2019) propôs utilizar a base *dicionário* juntamente com a hora do registro em cada base, assim, é feita uma integração temporal das duas bases.

Entretanto, a base de dados de rastreamento também registra pontos enquanto o veículo não está em serviço (deslocamentos para a garagem, trocas de linha, por exemplo) as quais não se deve atribuir uma linha ao registro. Para isso, Braga (2019) propôs a aplicação de um filtro espacial utilizando o itinerário das linhas presente no GTFS. Para cada linha do transporte público, foi criado um buffer de 400 metros em torno de todo o seu comprimento, definindo uma espécie de área de influência, onde os pontos fora dela são excluídos, como proposto por Braga (2019).

Essa abordagem é custosa do ponto de vista computacional, visto que processar grandes bancos de dados geoespaciais é altamente exigente em recursos computacionais (Gao, 2013). Na figura 7, é mostrado o filtro espacial aplicado por Braga (2019).

Figura 7 - Filtro espacial aplicado em cada linha por Braga (2019)



Fonte: Braga, 2019.

Liu (2005) em sua tese sustenta que uma abordagem baseada em soluções computacionais centralizadas, que é o tipo de abordagem disponível para este trabalho, é incapaz de lidar com o geoprocessamento de grandes bancos de dado de forma eficiente. Assim, visando diminuir a necessidade de utilizar tais ferramentas computacionalmente intensivas, defende-se que não é necessário aplicar o filtro espacial em todo o comprimento de cada linha, uma vez que, havendo validações no banco de bilhetagem para a linha em análise, basta verificar os pontos de início e de fim da linha para garantir que os registros compreendidos entre as duas localizações são, de fato, referentes a um período em que o veículo estava em serviço.

Além de validar os dados de GPS como de um veículo em serviço, nessa etapa também é identificado quando um veículo inicia e termina uma viagem, tornando possível o cálculo de métricas pertinentes, como o tempo de viagem ou a velocidade operacional para a linha. Uma abordagem muito similar pode ser usada quando se quer calcular indicadores não mais para uma linha inteira, mas para um certo trecho de um corredor, o qual é um dos objetivos deste trabalho. O filtro espacial simplificado proposto é mostrado na figura 8.

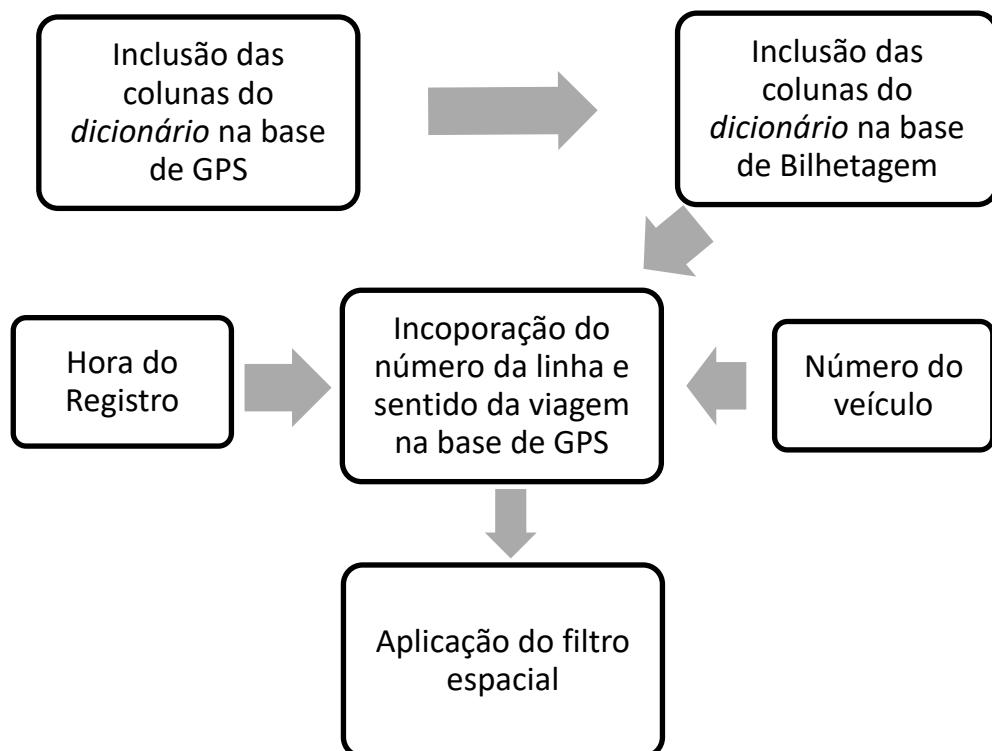
Figura 8 - Filtro Espacial Simplificado



Fonte: O Autor.

Os registros que não são localizados entre o início e fim da linha, são referentes a manobra dos veículos, deslocamento para a garagem, entre outros, e não são pertinentes à análise proposta nesse trabalho. Dessa forma, esses registros foram excluídos. O método usado por Braga (2019) para consolidar os dados é sumarizado na figura 8.

Figura 9 - Método de consolidação de Braga (2019)



Fonte: elaborado pelo autor

Ao final do processo proposto por Braga (2019), espera-se contar com cerca de 85% dos registros de localização dos veículos contendo também a informação da linha e do sentido da viagem. Propõe-se então, algumas etapas adicionais de forma a aumentar a proporção de registros contendo a informação da linha.

Primeiramente, separa-se na base de GPS os veículos com código *vehicleid* e os veículos com código *prefixo_carro* na base de bilhetagem de um certo dia que não constam na base *dicionário*. Os veículos na base de GPS serão usados juntamente com o itinerário das linhas na base GTFS para identificar à qual linha e sentido pertence os registros de GPS do veículo em questão. Uma vez identificada a linha, nos registros desse veículo é aplicado o método de Braga (2019) em ordem inversa para identificar o número *prefixo_carro* na base de bilhetagem correspondente ao *vehicleid* identificado na base de GPS, criando uma relação em ambos. Finalmente, o par *prefixo_carro* e *vehicleid* recém identificado é adicionado à base *dicionário* para que, em uma futura aparição do mesmo veículo, seja possível identificar sua respectiva linha diretamente pelo método de Braga (2019), evitando a utilização de um filtro espacial, pois, como é preciso verificar os registros do veículo para cada linha do SIT-FOR, essa etapa é muito intensiva computacionalmente.

Aos registros de GPS de cada veículo que não foi identificado a linha pelo método de Braga (2019) é aplicado um filtro espacial semelhante ao filtro espacial da etapa de validação, mas contendo todos os pontos de início e fim do itinerário das linhas contidas na base GTFS. Se esses registros são de um veículo em uso, naturalmente, haverá registros nos pontos de início e fim de uma ou mais linhas. Caso haja um único par início e fim de linha que satisfaça o filtro espacial, a linha é definida. Caso haja mais de uma linha possível para os registros, em possíveis casos de linhas que possuem seus pontos iniciais e finais inteiramente dentro do itinerário de outra linha (cenário bastante improvável, mas possível), uma etapa adicional é necessária

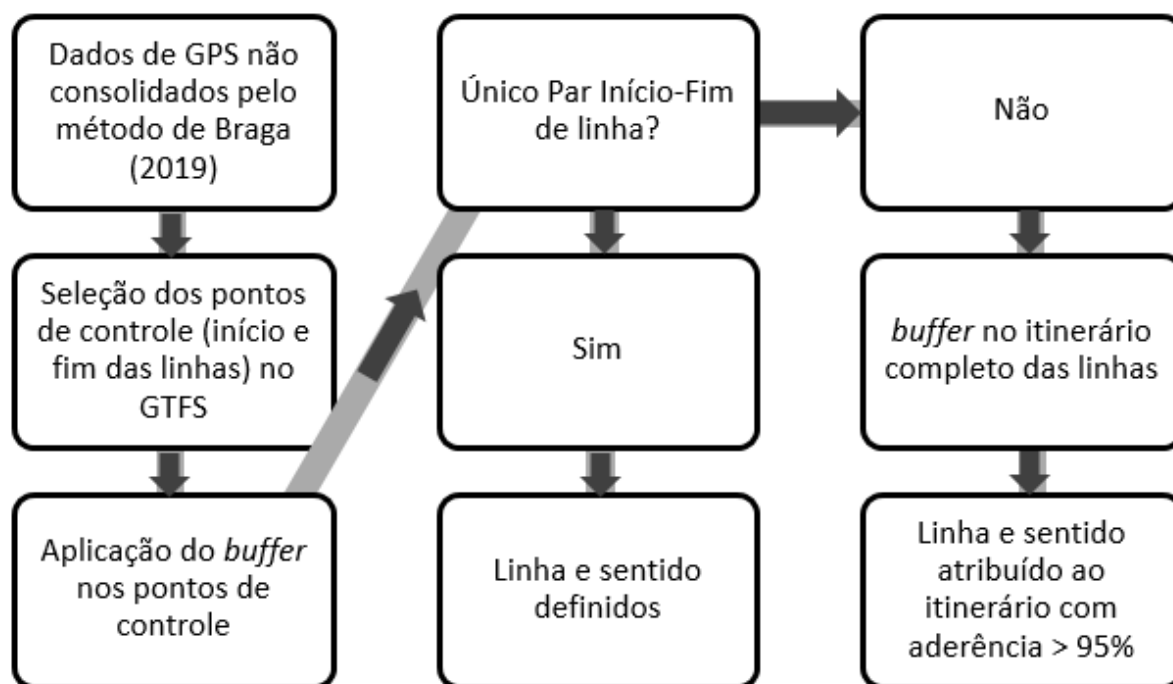
Nesse caso, é aplicado um filtro espacial a todo o itinerário das possíveis linhas a quem pertence os registros GPS, como no método de validação original de Braga (2015). O filtro espacial de uma das linhas possível terá maior aderência aos dados de GPS, a qual será atribuída aos dados. Nesse trabalho, adotou-se uma aderência mínima de 95% das observações, ou seja, 95% dos registros de GPS do veículo devem estar contidos na região que representa o itinerário da linha. Caso não a aderência seja inferior a 95%, os dados do veículo não serão atribuídos à nenhuma linha.

Vale ressaltar que, a tolerância de 5% dos dados fora do *buffer* do itinerário da linha foi estabelecido pela possibilidade de alguns registros estarem fora do *buffer* estabelecido por

algum motivo ligado ao equipamento de geolocalização. A aderência à 95% foi escolhida pois acredita-se que não haja linhas com itinerário 95% idêntico a outras. Ainda assim, devido ao modo como serão calculados os indicadores, uma hipotética linha 95% idêntica a outra ser considerada ou não à parte seria irrelevante para as conclusões deste trabalho.

O fluxograma do método adicional proposto é mostrado na figura 10.

Figura 10 - Fluxograma do método adicional de integração das bases



Fonte: O Autor.

Como o sucesso desse método depende da acurácia do itinerário das linhas na base de GTFS, foi verificado manualmente no arquivo se constavam todas as linhas que passam nos trechos de corredores estudados e se seus itinerários estavam corretos. Não foi encontrado erros. Ao final de todo o processo descrito nessa seção, os dados de GPS estão finalmente favoráveis ao cálculo dos indicadores.

5.2 Grupo de tratamento e grupo de controle

As faixas exclusivas foram implantadas de forma defasada, o que permite uma abordagem de um experimento natural. Na imagem 10 podemos ver que houve um rápido acréscimo na extensão das faixas exclusivas na cidade no período de maio de 2014 a dezembro de 2014, seguido de uma pausa até maio de 2015, quando foram retomadas novamente.

Figura 11 - Evolução das faixas exclusivas com indicação do período escolhido



Fonte: elaborado pelo autor.

Schiff *et al.* (2017) sustentam que se a escolha do grupo de tratamento e do grupo de controle é complementar aleatória, todos os outros fatores, exceto o tratamento (no caso, as faixas exclusivas), são cancelados entre os dois grupos, contanto que os grupos sejam suficientemente grandes, restando o efeito do tratamento, o qual é o interesse da análise. Contudo, as faixas exclusivas não foram implantadas todas, tampouco uma quantidade estatisticamente significativa, ao mesmo tempo. Isso torna impossível uma escolha totalmente aleatória dos grupos da análise.

Assim, na falta de grupos de controle e de tratamento aleatório, é necessário controlar para fatores observáveis que afetam a probabilidade de um corredor ter sido selecionado para receber faixa exclusiva ou não (Schiff *et al.*, 2017). Enquanto esses fatores observáveis são fáceis de controlar, pode haver também fatores não observáveis que afetam a probabilidade de escolha do grupo de tratamento, os quais são bem mais obscuros quanto ao controle, e certamente causam viés e podem levar a conclusões erradas. Ainda segundo Schiff *et al.* (2017), fatores não observáveis não ocorrem em estudos completamente aleatórios, mas podem ocorrer em estudos observacionais, que é o caso deste trabalho.

Assim, esses fatores deveriam ser cuidadosamente pensados no caso de estudos observacionais. Porém, identificou-se uma particularidade neste trabalho semelhante à do trabalho de Campbell e Brakewood (2017), visto na revisão da literatura, que permite diminuir consideravelmente a chance de fatores observáveis e não observáveis causarem viés na análise:

os corredores foram executados de forma defasada, pela mesma gestão da cidade. Isso nos deixa suficientemente confiantes de que a escolha do grupo de tratamento e do grupo de controle não carrega nenhum viés, desde que os corredores se restrinjam aos executados e previstos pela Prefeitura de Fortaleza, ou seja, todos fazem parte da população de corredores passíveis de receberem as faixas exclusivas.

Como visto na figura 11, uma grande extensão de faixas foi executada no ano de 2014, assim, escolheu-se as faixas exclusivas implantadas entre maio de 2014 e dezembro de 2014 como grupo de tratamento. Já as vias em que estava previsto implantação a partir de março de 2015 foram utilizadas no grupo de controle. Os meses escolhidos com bases na disponibilidade dos dados (não estava disponível dados anteriores a maio de 2014) e com o intuito de balancear a extensão de corredores nos dois grupos. Na tabela 3 é mostrado os corredores escolhidas para o grupo de tratamento, seguido das vias escolhidas para o grupo de controle na tabela 4 e é mostrado ambos os grupos espacialmente na imagem 12. Assim, foram 49,83 km de faixas exclusivas em 22 corredores no grupo de tratamento e 49,39 km de faixas exclusivas em 27 corredores no grupo de controle.

Tabela 3 - Corredores de tratamento

VIA	SENTIDO	TRECHO INÍCIO	TRECHO FINAL	DATA IMPLANTAÇÃO	d (km)
Avenida Dom Luís	Leste/Oeste	Rua Castro Monte	Rua Tibúrcio Cavalcante	07/06/2014	1,8
Avenida Santos Dumont	Oeste/Leste	Rua Nunes Valente	Rua Fonseca Lobo	07/06/2014	1,6
Avenida Carapinima	Norte/Sul	Avenida do Imperador	Avenida 13 de Maio	25/08/2014	0,45
Avenida Carapinima	Sul/Norte	Rua Padre Francisco Pita	Avenida Domingos Olímpio	25/08/2014	1
Avenida do Imperador	Norte/Sul	Avenida Duque de Caxias	Avenida Carapinima	25/08/2014	0,85
Avenida Tristão Gonçalves	Sul/Norte	Avenida Domingos Olímpio	Avenida Duque de Caxias	25/08/2014	0,56
Avenida da Universidade	Sul/Norte	Rua Padre Cícero	Avenida Domingos Olímpio	25/08/2014	1,53
Avenida Engenheiro Alberto Sá	Oeste/Leste	Rua Júlio Azevedo	Rua Prisco Bezerra	27/09/2014	0,69
Avenida Américo Barreira	Norte/Sul	Rua Monsenhor Furtado	Avenida Senador Fernandes Távora	29/09/2014	2,02
Avenida Américo Barreira	Sul/Norte	Avenida Senador Fernandes Távora	Rua Monsenhor Furtado	29/09/2014	2,01
Avenida Domingos Olímpio	Leste/Oeste	Avenida Aguanambi	Rua Justiniano de Serpa	29/09/2014	2,06
Avenida Domingos Olímpio	Oeste/Leste	Rua Justiniano de Serpa	Avenida Aguanambi	29/09/2014	2,08
Rua Justiniano de Serpa	Sul/Norte	Avenida Bezerra de Menezes	Avenida Domingos Olímpio	29/09/2014	0,2
Avenida José Bastos	Norte/Sul	Rua Padre Cícero	Rua Monsenhor Furtado	29/09/2014	2,13
Avenida José Bastos	Sul/Norte	Rua Monsenhor Furtado	Rua Porfírio Sampaio	29/09/2014	1,95
Avenida Antônio Sales	Oeste/Leste	Avenida Aguanambi	Rua Monsenhor Catão	31/10/2014	3,61

Avenida Bernardo Manuel	Norte/Sul	Rua Padre Ascenço Gago	Avenida Presidente Costa e Silva	31/10/2014	4,7
Avenida Bernardo Manuel	Sul/Norte	Avenida Presidente Costa e Silva	Rua Padre Ascenço Gago	31/10/2014	4,7
Avenida Jornalista Thomaz Coelho	Leste/Oeste	Rua Mucambinho	Avenida Presidente Costa e Silva	22/12/2014	1,97
Avenida Jornalista Thomaz Coelho	Oeste/Leste	Avenida Presidente Costa e Silva	Rua Florêncio Fontenele	22/12/2014	1,89
Avenida Presidente Costa e Silva	Leste/Oeste	Avenida Jornalista Thomaz Coelho	Avenida Godofredo Maciel	22/12/2014	6,01
Avenida Presidente Costa e Silva	Oeste/Leste	Avenida Godofredo Maciel	Avenida Jornalista Thomaz Coelho	22/12/2014	6,02

Fonte: elaborado pelo autor. Dados: PMF.

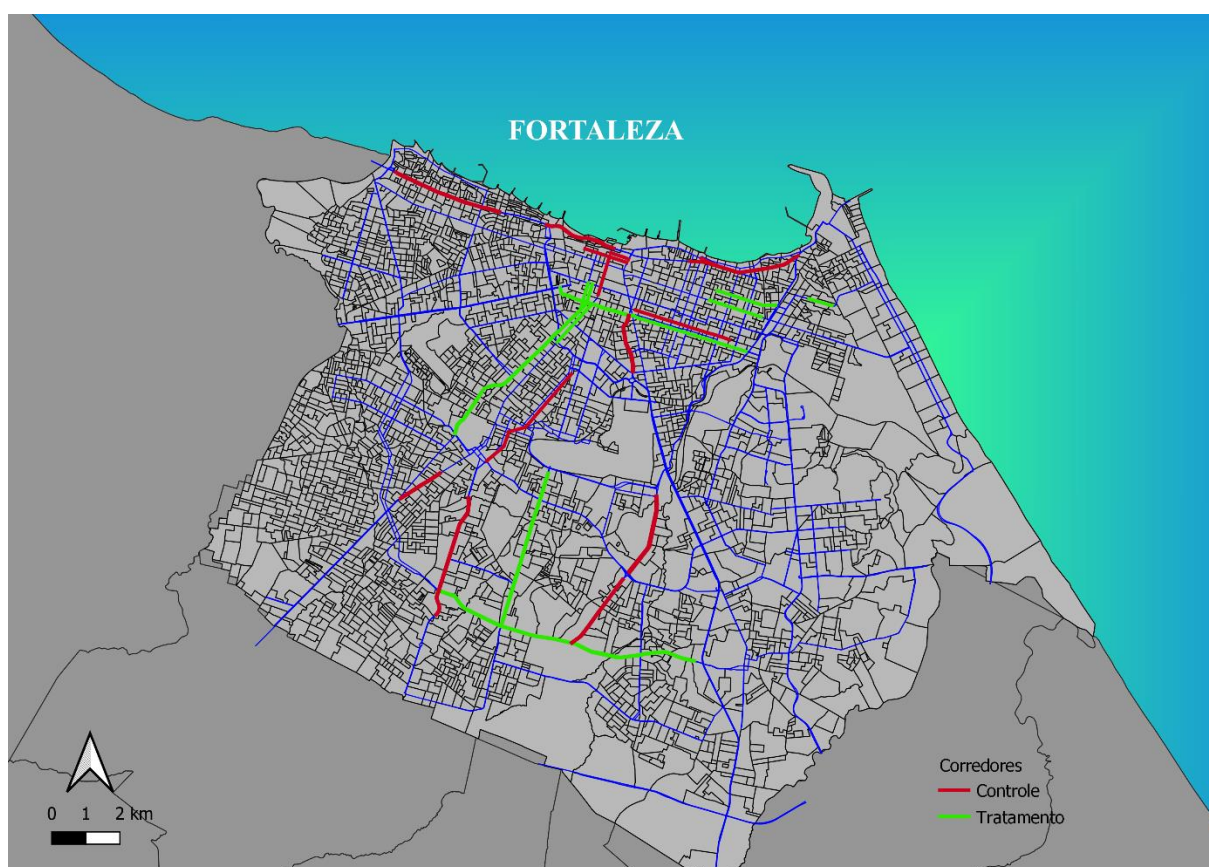
Tabela 4 - Corredores de controle

VIA	SENTIDO	TRECHO INÍCIO	TRECHO FINAL	DATA IMPLANTAÇÃO	d (km)
Avenida Professor Gomes de Matos	Sul/Norte	Rua Barão de Sobral	Rua Pré Nove	20/06/2015	2
Avenida da Abolição	Leste/Oeste	Rua Bauxita	Rua José Vilar	10/07/2015	2,53
Avenida da Abolição	Oeste/Leste	Avenida Barão de Studart	Avenida Almirante Henrique Sabóia	10/07/2015	2,73
Avenida Presidente Castelo Branco (1/2)	Leste/Oeste	Rua General Sampaio	Rua Jacinto Matos	13/07/2015	1,87
Avenida Presidente Castelo Branco (1/2)	Oeste/Leste	Rua Jacinto Matos	Rua General Sampaio	13/07/2015	2,2
Avenida Presidente Castelo Branco (2/2)	Leste/Oeste	Avenida Pasteur	Rua Ceci	13/07/2015	3,24
Avenida Presidente Castelo Branco (2/2)	Oeste/Leste	Rua Peri	Avenida Pasteur	13/07/2015	3,17
Avenida Alberto Craveiro	Norte/Sul	Rua Pedro Dantas	Avenida Doutor Silas Munguba	22/09/2015	2,31
Avenida Alberto Craveiro	Sul/Norte	Avenida Doutor Silas Munguba	Rua Pedro Dantas	22/09/2015	2,29
Rua Beni de Carvalho	Leste/Oeste	Rua Vicente Leite	Avenida Desembargador Moreira	27/11/2015	0,34
Rua Padre Valdevino	Leste/Oeste	Avenida Desembargador Moreira	Rua Rodrigues Júnior	27/11/2015	2,55
Avenida Monsenhor Tabosa	Oeste/Leste	Rua Francisco Virgílio de Vasconcelos	Avenida Barão de Studart	16/12/2015	0,57
Avenida Historiador Raimundo Girão	Leste/Oeste	Rua José Vilar	Avenida Rui Barbosa	16/12/2015	0,5
Avenida Osório de Paiva	Norte/Sul	Rua Osvaldo Aranha	Avenida Augusto dos Anjos	09/01/2016	1,26
Avenida Osório de Paiva	Sul/Norte	Avenida Augusto dos Anjos	Rua Osvaldo Aranha	09/01/2016	1,31
Avenida Presidente Juscelino Kubitschek	Norte/Sul	Rua Benvinda	Avenida Presidente Costa e Silva	13/03/2016	2,4
Avenida Presidente Juscelino Kubitschek	Sul/Norte	Avenida Presidente Costa e Silva	Rua Benvinda	13/03/2016	2,36
Rua César Rosas	Oeste/Leste	Rua Almirante Rubim	Rua Barão de Sobral	26/05/2016	0,49
Avenida Germano Franck	Sul/Norte	Avenida Doutor Silas Munguba	Rua Mundica Paula	28/09/2016	0,49
Rua Almirante Rubim	Sul/Norte	Rua Mundica Paula	Rua César Rosas	28/09/2016	0,63
Rua General Sampaio	Sul/Norte	Rua Meton de Alencar	Rua Castro e Silva	18/09/2017	1,2
Godofredo Maciel, Norte/Sul	Norte/Sul	R. Nereu Ramos	R. Arruda Câmara	27/03/2018	3,55

Godofredo Maciel, Sul/Norte	Sul/Norte	R. Arruda Câmara	R. Nereu Ramos	27/03/2018	3,55
Aguanambi N/S	Norte/Sul	Avenida Domingos Olímpio	Rotatória da Aguanambi	29/12/2018	2
Aguanambi S/N	Sul/Norte	Rotatória da Aguanambi	Avenida Domingos Olímpio	29/12/2018	2
João Moreira	Leste/Oeste	Av. Alberto Nepomuceno	R. General Sampaio	24/01/2019	0,55
Castro e Silva	Oeste/Leste	R. Padre Mororó	Av. Alberto Nepomuceno	24/01/2019	1,3

Fonte: elaborado pelo autor. Dados: PMF.

Figura 12 - Corredores de Tratamento e Controle



Fonte: elaborado pelo autor. Dados: PMF.

Para a escolha do período de análise é preciso que os corredores estejam em pleno funcionamento no mês escolhido antes das intervenções e no mês após as intervenções. Segundo o site de notícias FORTALBUS (2014) o primeiro dos corredores de tratamento (o binário Dom Luís/Santos Dumont) entrou em obras em junho de 2014, assim, escolheu-se o mês de maio de 2014 para o mês anterior ao tratamento.

Já para o grupo de controle, a avenida professor Gomes de Matos foi concluída em junho de 2015. Como a implantação de faixas exclusivas é um serviço rápido, pode-se usar o

mês de março de 2015 como mês de estudo após as intervenções. Dessa forma, os indicadores serão calculados para maio de 2014 e março de 2015.

Como se espera que as faixas exclusivas tenham maior efeito durante momentos de saturação dos corredores, os indicadores serão calculados para os 30 minutos mais carregados do pico da manhã em cada corredor. Os finais de semana seriam dias de menor saturação dos corredores, contudo, não foram excluídos da análise que desconsidera volume nos corredores, pois ambos os meses analisados possuem o mesmo número de finais de semana, cinco, o que cancelaria seu efeito ao aplicar o método de inferência causal. Quanto a feriados, excluiu-se o dia 1º maio de 2014 e os dias 19 e 25 de março de 2015 (feriados estaduais). Entende-se que os horários de pico podem variar de corredor a corredor, a depender da localização na cidade. Mas, como visto na figura 12, há uma boa distribuição espacial dos corredores nos dois grupos da análise, então espera-se que essas diferenças sejam canceladas.

Para definir a hora de pico se dispõe dos dados de fiscalização eletrônica, que registram a quantidade de veículos que passam pela via. Isso permite definir a hora da manhã e a hora da tarde com o maior número de veículos, o volume, para cada hora do dia. Uma amostra dos dados de fiscalização eletrônica é mostrada na tabela 5. A informação referente à placa do veículo é criptografada.

Tabela – 5 Amostra de dados de fiscalização Eletrônica

	Equip	Data	Hora	Faixa	Vel Perm	Vel Medida	Tamanho	Perfil	Placa
1	AT007	01/03/2015	00:00:09	2	60	48	1.7	1	#####
2	AT007	01/03/2015	00:00:31	2	60	54	4	2	#####
3	AT007	01/03/2015	00:00:42	1	60	62	3.8	2	#####
4	AT007	01/03/2015	00:01:31	2	60	54	3.8	2	#####
5	AT007	01/03/2015	00:01:33	1	60	56	3.7	2	#####

Fonte: O autor

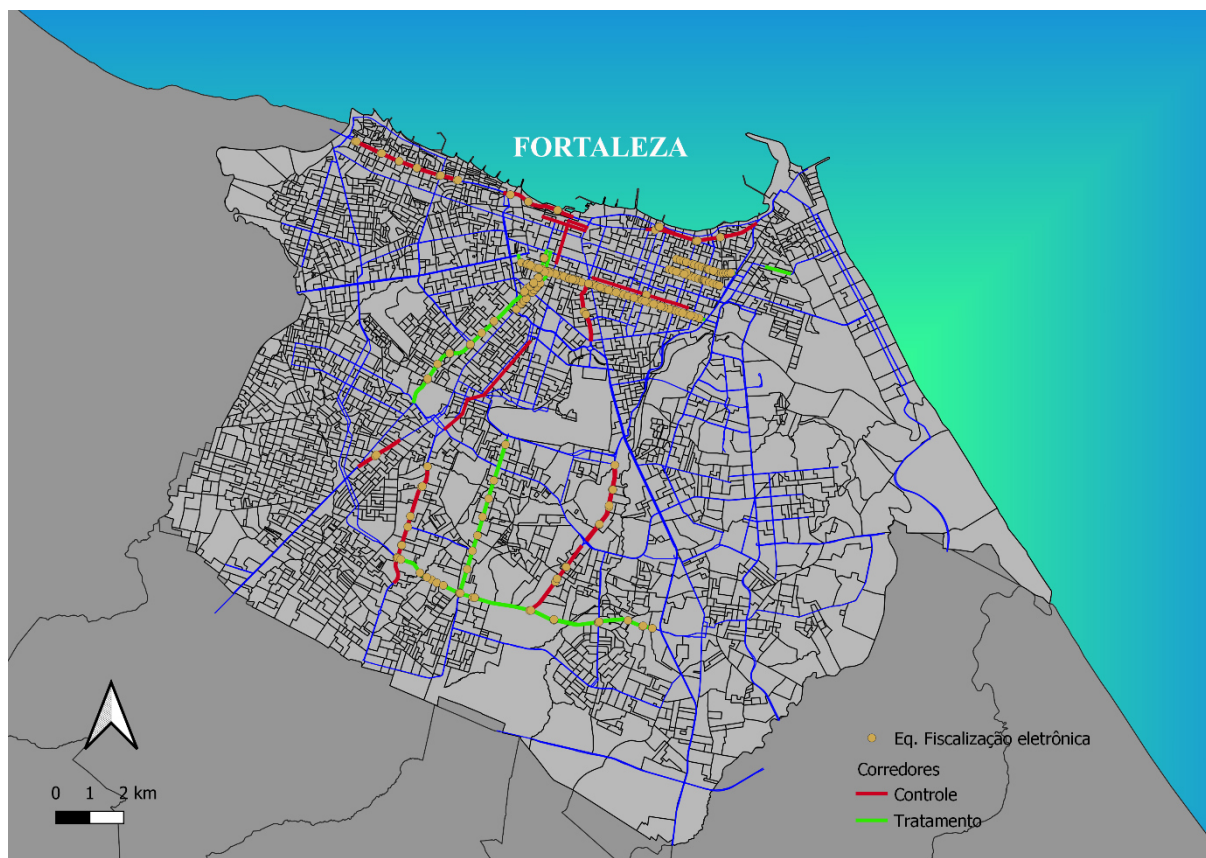
Foi feito histograma para cada hora do dia no mês para identificar visualmente os 30 minutos mais carregados do pico da manhã.

Contudo, ao calcular os volumes horários foi excluído os finais de semana e feriados, 1 de maio de 2014 e 25 de março de 2015 (somente para a segunda análise, pois pretende-se fazer um modelo de controle utilizando um modelo de regressão a partir do volume). A partir de inspeção visual, escolheu-se um período de 30 minutos durante o maior volume da manhã. Ressalta-se que se optou por um intervalo de apenas 30 minutos devido à grande exigência computacional para processamento dos dados.

Por exemplo, verificou-se a partir de uma rápida análise dos dados que finais de semana possuíam maior volume em altas horas da noite em relação a dias úteis, o que era

esperado. Como não se sabe se o comportamento do fenômeno varia diferentemente entre feriados e entre finais de semana, optou-se por excluí-los no segundo modelo, visto que este é sensível à magnitude dos valores de volume. Esse método de controle será descrito em mais detalhes na seção 5.3. A localização dos equipamentos de fiscalização eletrônica para cada corredor em estudo é mostrada na figura 13.

Figura 13 - Localização dos equipamentos de fiscalização eletrônica



Fonte. O Autor. Dados: PMF.

5.3 Indicadores

A partir da revisão bibliográfica foram definidos os indicadores. Como o método de inferencial causal visa justamente identificar diferenças devido a um tratamento em um grupo, no caso, as faixas exclusivas, o valor absoluto dos indicadores não é o mais importante, e sim a diferença entre os dois cenários.

Fez-se uma análise descritiva dos indicadores abaixo para os meses de maio de 2014 e março de 2015.

Os indicadores, bem como a justificativa para sua escolha e sua definição são os seguintes:

- **Velocidade Média Operacional:** é a velocidade média dos veículos do transporte público excluindo os tempos de manobra nos terminais. Foi calculado por trecho. A velocidade operacional é um indicador que, segundo a literatura, deve ser impactado pelas faixas exclusivas (Arbex e Da Cunha, 2016). Para calcular esse indicador para o corredor, separou-se os dados por veículo e verificou-se o tempo que o veículo precisou para ir do início do trecho ao fim do trecho. A formulação do indicador para linhas é a seguinte:

$$V_{op(l)} = \frac{\sum_{i=1}^n \frac{D_i}{t_i}}{n}$$

Onde:

- $V_{op,l}$ é a velocidade média operacional das linhas do grupo de controle ou do grupo de tratamento;

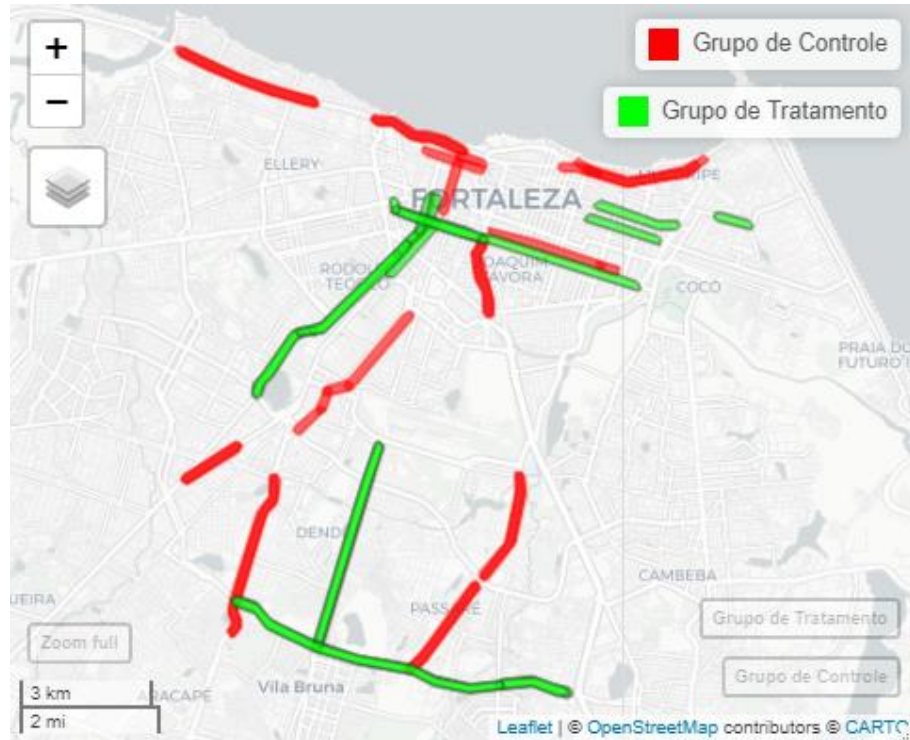
- f é uma variável binária: 0 para grupo de controle e 1 para grupo de tratamento;

- z é o número de linhas no grupo f ;

Para o cálculo da velocidade média, foi feito um *buffer* ao longo dos trechos selecionados. Os dados de GPS foram filtrados para esse *buffer* e depois foi identificado o momento do primeiro e do último registro no *buffer*. A distância entre os registros é calculada como uma linha reta na superfície curva da terra.

Os *buffers* utilizados são mostrados na figura 14.

Figura 14 - *Buffers* de controle e de tratamento



Fonte: O autor.

Após aplicado o filtro espacial, são mantidos apenas os registros dentro do *buffer*, como mostrado na figura 15.

Figura 15 - Registros de GPS dentro dos *Buffers*



Fonte: O autor.

- **Variância da velocidade média operacional:** é a variância da velocidade média operacional de cada grupo por trecho. Espera-se que as faixas exclusivas reduzam a variação da velocidade operacional, e, por conseguinte, a variância. Assim, foram calculados:

 - $s_{c0,k}$: é a variância amostral das velocidades médias operacionais dos veículos no trecho de controle k em 2014.
 - $s_{c1,k}$: é a variância amostral das velocidades médias operacionais dos veículos no trecho de controle k em 2015.
 - $s^2_{c0,j}$: é a variância amostral das velocidades médias operacionais dos veículos no trecho de tratamento j em 2014.
 - $s^2_{c1,j}$: é a variância amostral das velocidades médias operacionais dos veículos no trecho de tratamento j em 2015. Em seguida, a variância das velocidades de cada veículo no corredor i foram somados. Entende-se que a soma das variâncias das velocidades de veículos diferentes incorporará variações intrínsecas das linhas as quais eles servem, ou outros fatores próprios de cada veículo, à soma das variâncias. Contudo, como será feito a diferença das diferenças deste indicador, espera-se que essa variância própria de cada veículo seja eliminada. Além disso, para a validade dessa operação, é tomado como premissa de que as velocidades médias dos diferentes veículos são independentes.
- **Variância dos *headways*:** é a variância do tempo entre veículos de uma mesma linha nos corredores. Espera-se que as faixas exclusivas diminuam a variação do *headway* da linha. Esse indicador foi calculado da seguinte maneira:

 - *headways* dos veículos das linhas por corredor: fixou-se um ponto final do corredor e gravou-se o horário em que cada veículo servindo uma linha específica passou por esse ponto. O *headway* é a diferença entre os horários. Foi calculado para os corredores do grupo de controle ($S^2_{hc,0}$) e de tratamento ($S^2_{hc,1}$), para o ano de 2014 (0) e de 2015 (1). Em seguida, a variância dos *headways* de cada linha no corredor i foram somados. Entende-se que a soma das variâncias dos *headways* de linhas diferentes incorporarão variações intrínsecas de cada linha à soma das variâncias. Contudo, como será feito a diferença das diferenças deste indicador, espera-se que essa variância própria de cada linha seja eliminada. Além disso, para

a validade dessa operação, é tomado como premissa de que os *headways* das diferentes linhas são independentes.

5.4 O modelo de estimação do impacto das faixas exclusivas

A execução defasada das faixas exclusivas permite uma abordagem de quase experimento. Como visto na revisão da literatura, o método da diferença nas diferenças é o candidato principal a estimar o impacto das faixas exclusivas (Schiff *et al.*, 2017) nos indicadores definidos na seção 5.3.

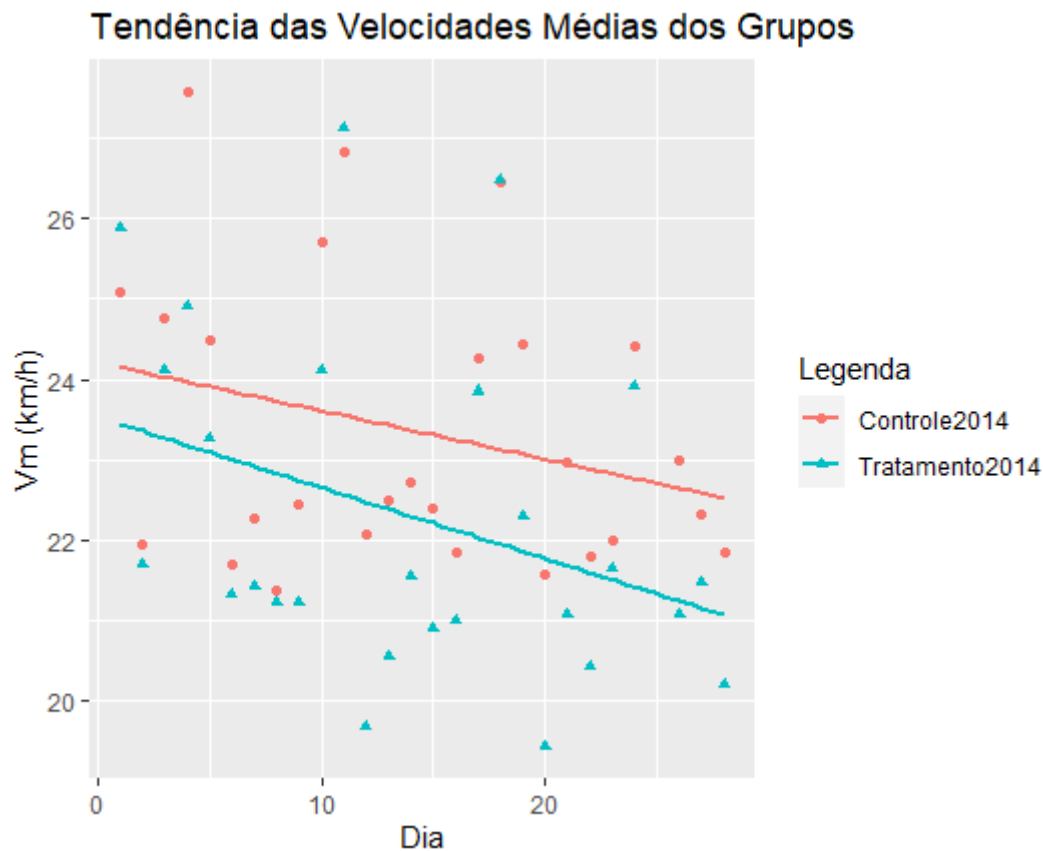
O método da diferença nas diferenças pode ser usado quando há dados anteriores e após uma intervenção, estimando o impacto da intervenção. O método consiste em comparar a diferença entre os indicadores do grupo que recebeu o tratamento e o grupo de controle anteriormente ao tratamento e a mesma diferença após a implementação da intervenção.

Segundo Khandker *et al* (2010), as principais premissas para o uso do método são que potenciais variáveis não observáveis são constantes no tempo ou possuem o mesmo efeito nos dois grupos: tratamento e controle. Assim, a premissa dos caminhos paralelos se mantém e o grupo de controle atua como estimativa do contrafactual, que não pode ser observado (Mora e Reggio, 2012). Dessa forma, a diferença entre os dois caminhos paralelos é a estimativa do impacto da intervenção.

A utilização de dados de intervenções defasadas no tempo para separar os grupos de controle e de tratamento por si só adiciona robustez a essas duas principais premissas da diferença nas diferenças (Campbell e Brakewood, 2017). Ademais, a presença de corredores de diferentes pontos da cidade, de diferentes usos de solo e características geométricas adiciona confiança de que essas premissas são válidas. Potenciais variáveis que podem não se manter constante, tampouco afetar igualmente os dois grupos durante o período do estudo são discutidas na seção 5.5, juntamente com um método de controle e a justificativa para não serem incluídas no estudo.

Para aumentar a confiança de que os dois grupos seguem a premissa dos caminhos paralelos, foi plotado um gráfico das velocidades operacionais médias dos dois grupos de maio de 2014 e março de 2015 para verificar visualmente se ambos os grupos possuíam tendências similares. O referido gráfico é mostrado a figura 16.

Figura 16 - Tendência das velocidades médias operacionais dos dois grupos



Fonte: o Autor.

Pelo gráfico da figura 16, pode-se ver que o grupo de tratamento e o grupo de controle possui tendências muito similares para 2014, corroborando a escolha dos referidos grupos.

A formulação do método é apresentada para cada indicador nas subseções que seguem.

5.4.1 Estimativa do impacto na Velocidade Média Operacional

Seja $V_{opl,1}$ a velocidade operacional dos veículos nas linhas do grupo de tratamento e $V_{opl,0}$ a velocidade operacional dos veículos nas linhas do grupo de controle. Considerando $t=0$ o mês anterior a execução das faixas exclusivas no grupo de tratamento (maio/2014) e $t=1$ o período após a implementação das faixas (março/2015), o impacto das faixas exclusivas no indicador pode ser estimado por:

$$(E[V_{opl,1}|t = 1] - E[V_{opl,1}|t = 0]) - (E[V_{opl,0}|t = 1] - E[V_{opl,0}|t = 0]) = \beta_l$$

Onde β_l é o efeito causal estimado médio da implantação das faixas exclusivas na velocidade operacional dos veículos das linhas impactadas pela intervenção.

5.4.2 *Estimativa do impacto na variância da Velocidade Média Operacional*

Seja $S_{c1,j}$ a variância amostral das velocidades médias operacionais dos veículos no trecho de tratamento j e $S_{c0,k}$ a variância amostral das velocidades médias operacionais dos veículos no trecho de controle k , o impacto no desvio padrão dessas velocidades é estimado por:

$$(E[S_{c1,j}|t = 1] - E[S_{c1,j}|t = 0]) - (E[S_{c0,k}|t = 1] - E[S_{c0,k}|t = 0]) = \alpha_c$$

Onde α_c é o efeito causal estimado médio da implantação das faixas exclusivas no desvio padrão da velocidade operacional dos veículos nos trechos onde fora implantado faixas exclusivas.

O valor esperado de cada variável corresponde à média da soma das variâncias no respectivo grupo.

5.4.3 *Estimativa do impacto na variância do headway*

Seja S_{L1} a variância amostral dos headways dos veículos nos trechos de tratamento e S_{L0} a variância amostral dos headways dos veículos dos trechos de controle. Para o mesmo intervalo de tempo usado no indicador anterior, o impacto das faixas exclusivas no desvio padrão das velocidades operacionais na linha pode ser estimado por:

$$([S_{L1}|t = 1] - [S_{L1}|t = 0]) - ([S_{L0}|t = 1] - [S_{L0}|t = 0]) = h_c$$

Onde h_c é o efeito causal estimado médio da implantação das faixas exclusivas na variância do headway dos veículos nos corredores onde fora implantado faixas exclusivas.

5.4.4 *Teste de hipótese dos indicadores*

Essa etapa visa verificar se o impacto em cada indicador é significativo estatisticamente. Para isso, fez-se testes de hipótese para cada indicador. As hipóteses alternativas para cada indicador são apresentadas na tabela 5.

Tabela 5 - Teste de hipótese dos impactos nos indicadores

Impacto no indicador	Hipótese alternativa (H_a)	Valor-p
β_c	$H_a = \beta_c > 0$	$p < 0,05$
α_c	$H_a = \alpha_c < 0$	$p < 0,05$
h_c	$H_a = h_c < 0$	$p < 0,05$

Os cálculos estatísticos foram feitos no software *RStudio* utilizando linguagem de programação *R* e diversos pacotes estatísticos. O teste de hipótese para a diferença de velocidade média β_c é o teste-t não pareado. Neste caso, tem-se a premissa de que as velocidades médias do grupo de controle e do grupo de tratamento não são relacionadas. Como as amostras são muito superiores à 30 observações, o teste-t pode ser usado.

5.5 Controle de variáveis, modelo placebo e considerações finais

Para explorar a possibilidade de a existência das faixas exclusivas terem alterado a distribuição do volume apenas em um dos grupos, ou ainda de a escolha dos corredores para receberem faixas exclusivas ter algum motivo implícito na saturação da via, o que resultaria em um viés, propõe-se um modelo de diferença nas diferenças alternativo com controle do volume nos corredores. Por exemplo, uso do solo, baixas velocidades operacionais, nível alto de saturação das vias, diferenças sociais e econômicas das regiões beneficiadas, características geométricas dos corredores, distância entre paradas, presença de binários, quantidade de interseções e etc.

Mesmo que algum motivo não observado esteja por trás da escolha dos corredores, o fato de que para os corredores usados para controle também estava previsto implantação de faixas exclusivas traz confiança de que as últimas variáveis citadas possam ser canceladas pelo método da diferença nas diferenças. No caso de binários, uma variável binária poderia ser adicionada ao modelo proposto em 6.4 para isolar seu efeito. Porém, como não é objetivo desse trabalho isolar os efeitos dos binários, a variável binária não foi incluída, visto que se espera que o efeito decorrente dos binários seja cancelado pela diferença das diferenças. Todavia, possíveis impactos devido ao volume podem ocorrer de forma diferente em cada grupo, tornando necessário controlar o modelo para esse caso.

É proposto um modelo de regressão da velocidade operacional dos veículos nos trechos de corredores de tratamento e de controle, formulado da seguinte maneira:

$$V_{op,f} = \beta_f + \alpha_f T_{f,t} + \theta \frac{\sum_{i=1}^n V_{i,f}}{n} + \varepsilon$$

Onde:

- $V_{op,f}$ é a velocidade operacional média do grupo f ;
- $V_{i,f}$ é o volume na hora de pico no corredor i ;
- f é uma variável binária: 0 para grupo de controle e 1 para grupo de tratamento;
- ε é o erro residual;
- α é o parâmetro da regressão que é o objeto de interesse do modelo.
- θ é o parâmetro da regressão referente ao volume no corredor.

O modelo foi calculado usando os dados do mês de maio de 2014, para o grupo de controle e para o grupo de tratamento e usando os dados de março de 2015 para os mesmos dois grupos. Considerando $t=0$ o período antes da implantação das faixas e $t=1$ o período após, pelo método da diferença nas diferenças, o impacto da variação do volume na velocidade operacional são os seguintes:

$$([\alpha_1|t = 1] - E[\alpha_1|t = 0]) - ([\alpha_0|t = 1] - [\alpha_0|t = 0]) = \alpha_d$$

Onde α_d é o efeito causal estimado médio da variação do volume causado pelas faixas exclusivas no coeficiente angular da reta de regressão.

E:

$$([\beta_1|t = 1] - E[\beta_1|t = 0]) - ([\beta_0|t = 1] - [\beta_0|t = 0]) = \beta_d$$

Onde β_d é o efeito causal estimado médio da variação do volume causado pelas faixas exclusivas no coeficiente angular da reta de regressão.

Como no modelo anterior, é verificado se as diferenças encontradas são estatisticamente significativas, utilizando teste de hipótese com valor- $p < 0,05$. As hipóteses alternativas são:

1. $H_a = \alpha_d < 0$
2. $H_a = \beta_d \neq 0$

Além disso, foi conduzido um experimento placebo para testar que ambos os grupos seguem a mesma tendência. Para isso, realizou-se o procedimento de estimativa do impacto na velocidade operacional descrito anteriormente, somente para os corredores, para os meses de março de 2018 e outubro de 2018, quando todos os corredores já possuíam faixas exclusivas. É esperado efeito nulo nesse cenário.

Por fim, reconhece-se que nem todas as variáveis podem ter sido canceladas, principalmente quanto ao número de interseções e distância entre paradas. Assim, é premissa nesse trabalho que essas variáveis não sofrem variação após as intervenções.

6 IMPACTO DAS FAIXAS EXCLUSIVAS NOS INDICADORES DE DESEMPENHO DO SIT-FOR

Esta seção apresentará os resultados obtidos ao aplicar o método descrito na seção 5 nos dados de GPS do SIT-FOR. A análise começou pela definição da hora de pico, através dos dados de volume da fiscalização eletrônica, a partir da qual selecionou-se um período de 30 minutos mais carregado no horário da manhã. Em seguida, calculou-se a velocidade média dos veículos nos corredores do grupo de controle e de tratamento para os cenários *ante* intervenção e *post* intervenção. Uma análise descritiva da velocidade média desses grupos é apresentada.

Embora a integração das bases de dados seja necessária para a análise do indicador variância do *headway* proposto, devido ao grande consumo de poder computacional para o cálculo da velocidade média, bem como diversos empecilhos no seu cálculo (que serão detalhados mais a frente), não foi possível implementar o método previsto para integração das bases e cálculo da variância do *headway* de forma plena.

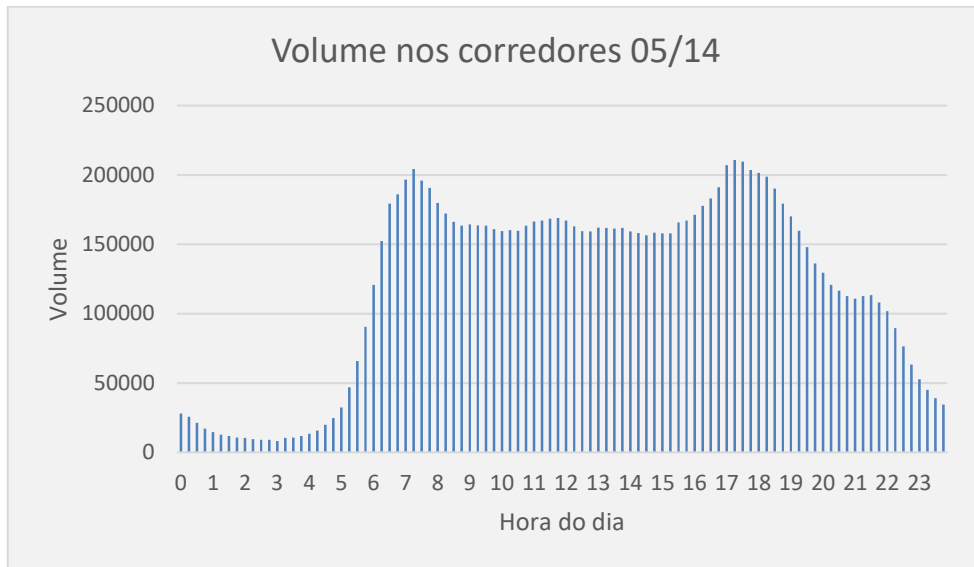
Com o indicador de velocidade média calculado para os dois cenários e os dois grupos, foi aplicado o método da diferença nas diferenças como descrito na seção 6.3 deste trabalho, seguido do teste de significância t de Welch. O processo foi repetido para a variância das velocidades médias. Uma interpretação e discussão do autor sobre os resultados é oferecida.

6.1 Determinação do período diário de análise

Os equipamentos de fiscalização eletrônica foram filtrados para os corredores do grupo de controle e para o grupo de tratamento utilizando o *software* QGIS. De posse do código dos equipamentos para cada grupo, os volumes foram agregados a cada 15 minutos, montando os histogramas.

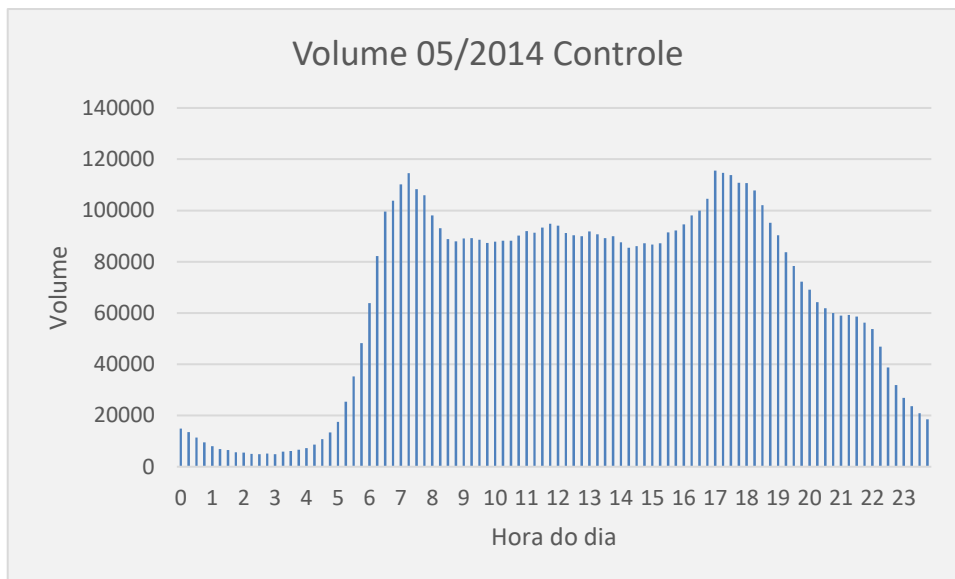
Os histogramas para o mês de maio de 2014 são mostrados nas figuras 17, 18 e 19 bem como os histogramas para o mês de março de 2015 são apresentados nas figuras 20, 21 e 22.

Figura 17 - Volume mensal nos corredores em maio de 2014



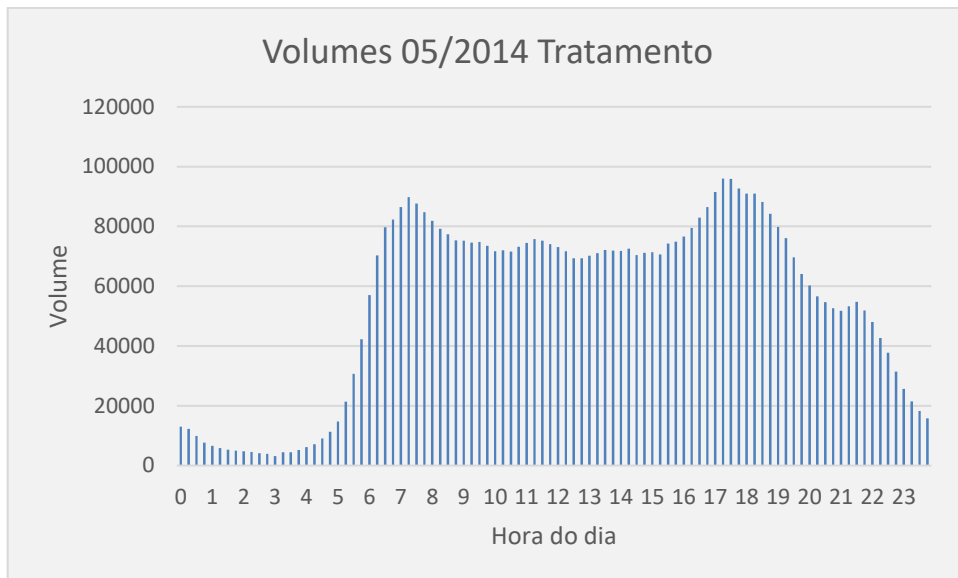
Fonte: O autor. Dados: PMF.

Figura 18 - Volume mensal nos corredores do grupo de controle em maio de 2014



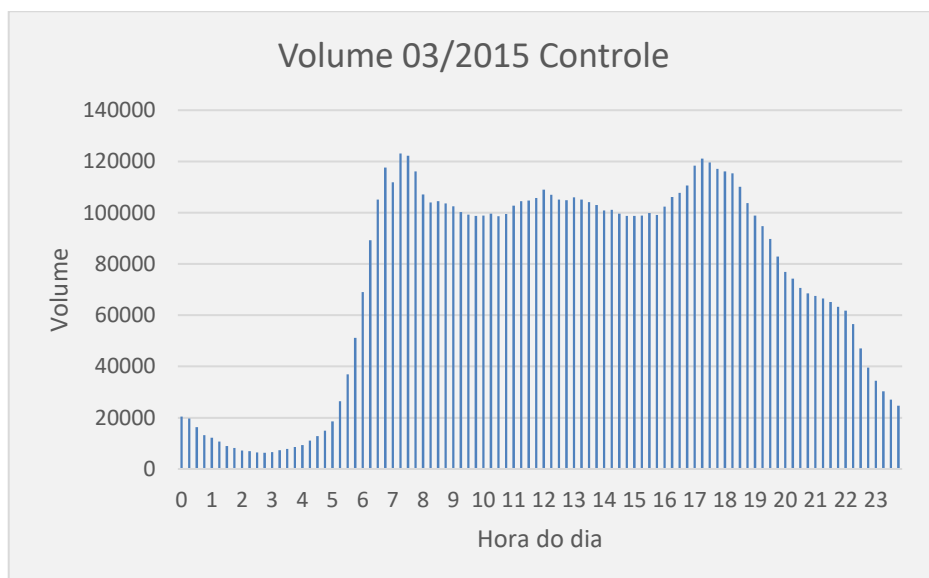
Fonte: O autor. Dados: PMF.

Figura 19 - Volume mensal nos corredores do grupo de tratamento em maio de 2014



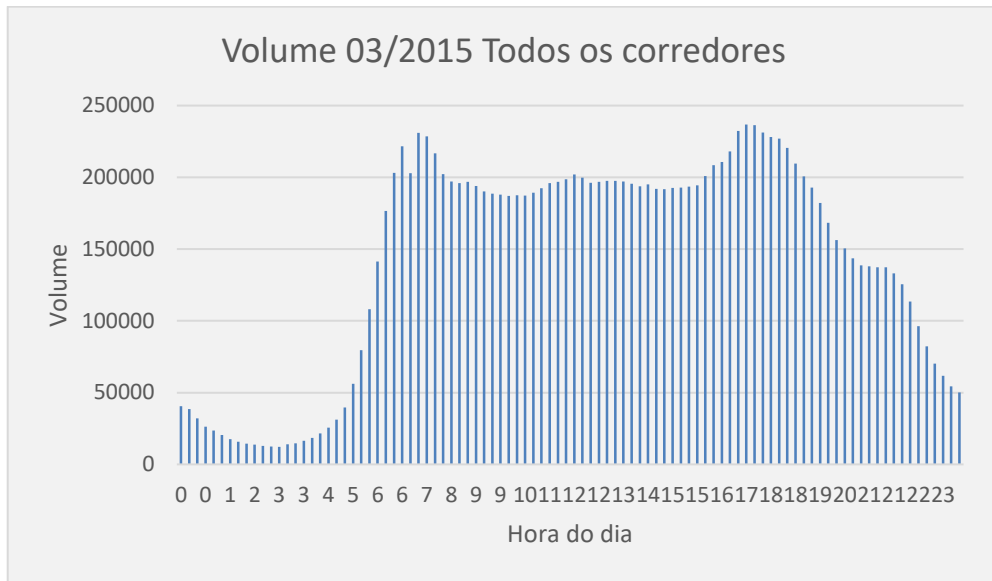
Fonte: O autor. Dados: PMF.

Figura 20 - Volume mensal nos corredores do grupo de controle em março de 2015



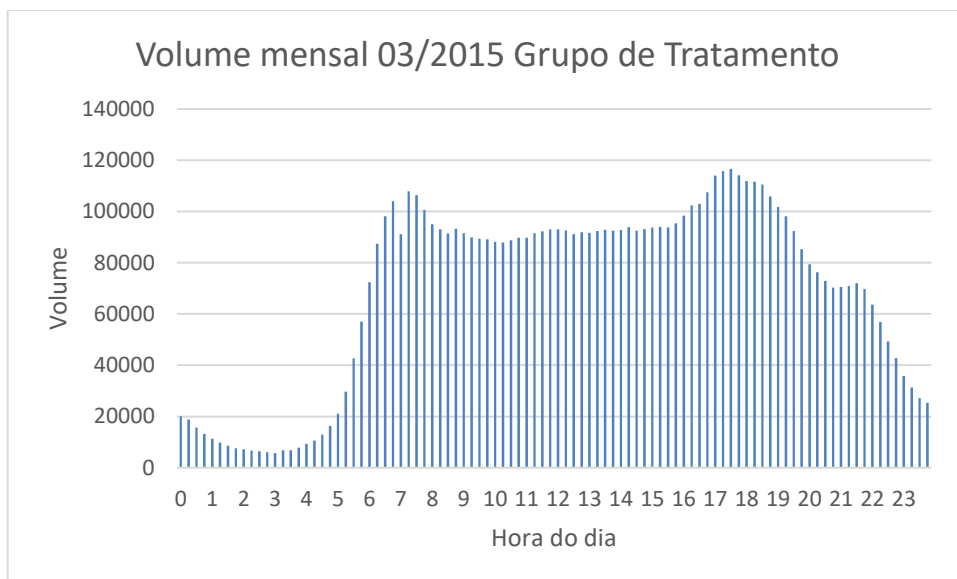
Fonte: O autor. Dados: PMF.

Figura 21 - Volume mensal nos corredores em março de 2015



Fonte: O autor. Dados: PMF.

Figura 22 - Volume mensal nos corredores do grupo de tratamento em março de 2015



Fonte: O autor. Dados: PMF.

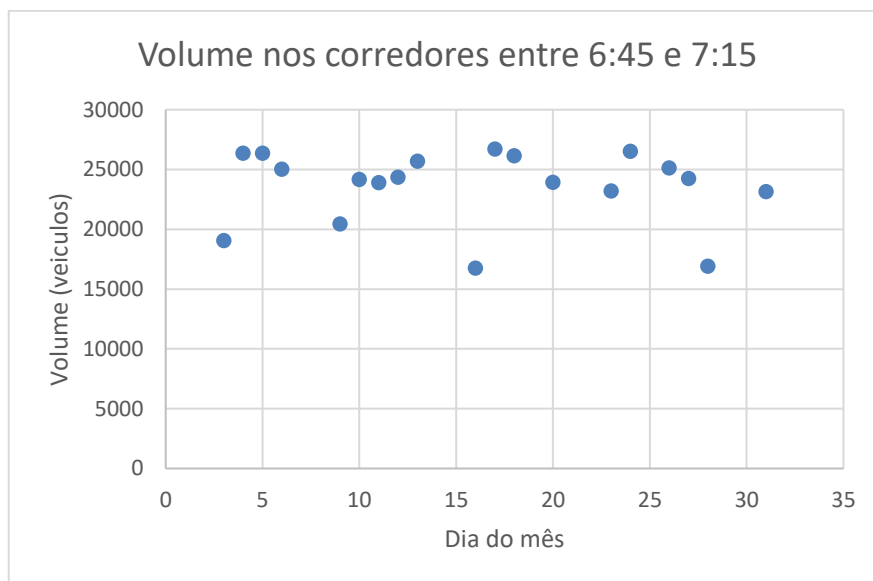
Através de inspeção visual nos histogramas dos volumes nos corredores é possível constatar que não há diferença perceptível quanto à forma entre os histogramas dos grupos de controle e tratamento, tampouco entre o ano de 2014 e 2014. Em todos os casos, o pico da manhã ocorre entre 6:30 e 8:00. Como já mencionado, por motivo de limitação computacional, foi selecionado o intervalo de 30 minutos entre 6:45 e 7:15.

A similaridade nos histogramas de volume também corrobora a hipótese de que os grupos de controle e tratamento são similares, bem como estão bem distribuídos na cidade. Entretanto, o grupo de controle possui volume consistentemente superior ao volume no grupo de tratamento. Além da possibilidade de os corredores do grupo de controle terem realmente um maior volume, também é uma possibilidade o fato de o grupo de controle possuir mais corredores.

Para utilizar os dados de volume no modelo de regressão linear com DiD, foi necessário selecionar os reagrupar os dados de volume para cada 30 minutos, de 6:45 à 7:15 da manhã.

Os volumes nos corredores no intervalo de estudo para cada dia do mês são mostrados na figura 23.

Figura 23 - Volume nos corredores entre 6:45 e 7:15 para o mês de março de 2015



Fonte: O Autor. Dados: PMF.

6.2 Cálculo dos indicadores

O cálculo da velocidade média nos trechos foi notadamente uma parte trabalhosa. A forma como os dados estão armazenados e a forma como se tentou calcular a velocidade média causaram contra-tempos. A saber, o tamanho do buffer utilizado precisou ser ajustado por meio de tentativas, visto que buffers grandes demais resultavam em registros de GPS sendo filtrados para dois buffers de corredores ao mesmo tempo, enquanto que buffers muito pequenos não eram sempre capazes de capturar os registros de GPS no início e fim do buffer do corredor.

Isso se deve ao fato de que os veículos transmitem sua posição a cada 30s, ou seja, tempo suficiente para o veículo percorrer uma distância considerável e escapar do buffer.

Após várias tentativas com buffers diferentes, chegou-se a um valor razoável de 100 m, o qual não gerava registros atribuídos à mais de um buffer e gerava uma quantidade razoável de registros capturados no início e no final do buffer, gerando os pares de localização usados para calcular a velocidade média.

A distância entre o par de pontos de localização utilizados foi a distância reta na superfície curva da terra. Entende-se que isto pode resultar em distâncias menores que os reais, e, por conseguinte, velocidades médias maiores que os reais. Contudo, como é possível observar na figura 12, os trechos de corredores são majoritariamente retos, tornando a diferença entre a distância em linha reta e a distância real em malha pequena. Ressalta-se que não foi utilizado a distância em rede usando a *API* da *google* devido à grande quantidade de registros, que excedia em grande escala o disponibilizado gratuitamente pela *google*. Embora visualmente tenha-se a impressão que os trechos são majoritariamente retos, seria interessante ter sido feito uma comparação entre as distâncias em linha reta nos trechos e a distância em malha viária dada pela *API* da *google*.

Outra dificuldade surgiu devido aos veículos que tiveram os registros capturados somente no início ou somente no final do buffer, os quais ficavam sem par de localização final e inicial, e devido a forma como o algoritmo foi pensado, isso causaria erro na ordenação dos pares corretos de localização. Para contornar esse problema, os dados foram filtrados três vezes. Inicialmente verificou-se os registros que apareciam apenas uma vez no buffer, e estes foram descartados. Segundamente, foi verificado os registros que apareciam três vezes no buffer do corredor, isto é, veículos que completavam uma viagem e chagavam a iniciar a segunda viagem dentro do intervalo de 30 minutos. O registro da viagem incompleta desses veículos foi também excluído. Por fim, os registros foram novamente verificados quanto à presença de pares de localização defeituosos, que também foram excluídos.

Isto traz à tona outro aspecto do cálculo da velocidade média. Devido à falta de identificação das viagens, apenas uma viagem por veículo poderia ser reconhecida no intervalo escolhido. Esse foi um dos motivos, aliado à grande necessidade computacional, da definição do intervalo de 30 minutos para análise. Verificou-se que esse tempo é inferior ao ciclo dos veículos, de forma a cada veículo realizar apenas uma viagem completa no trecho.

Além do supradito, foram encontradas velocidades muito baixas (da ordem de 1km/h), que fizeram com que o autor suspeitasse de que poderia haver algum erro em alguns poucos registros do *dataset*. Como não foi possível fazer um código para ter certeza de que

todas as velocidades calculadas eram legítimas, o autor verificou manualmente alguns tempos de viagem e velocidades que pareciam baixos demais. Todos os casos de velocidade média baixa verificados manualmente se mostraram legítimos, pois os veículos selecionados foram verificados utilizando todos os registros, à cada 15s, e constatou-se que realmente o algoritmo estava correto. Contudo, vale ressaltar que devido à grande quantidade, não se pôde verificar todos os veículos com velocidade muito baixa manualmente, tampouco é possível estar totalmente seguro que o código feito funciona em 100% dos dias analisados, devido à natureza imprevisível da organização dos dados de GPS.

Uma amostra dos dados de velocidade média para cada grupo e cada ano é mostrado nas tabelas 6, 7, 8 e 9.

Tabela 6 - Amostra de indicadores de velocidade média calculados para o grupo de controle em maio de 2014

Longitude	Latitude	id	Dist (m)	Data-Hora	Logradouro	diff Temp (s)	Vm (km/h)
-38,5308	-3,7227	2025	479	07/05/2014 06:58	Rua Castro e Silva	103	16,74
-38,5258	-3,7417	2031	1442	07/05/2014 07:01	Av, Aguanambi 2	244	21,27
-38,5444	-3,7149	2061	302	07/05/2014 07:11	Av, Presidente Castelo Branco 4	49	22,20
-38,5606	-3,7106	2063	301	07/05/2014 07:13	Av, Presidente Castelo Branco 2	62	17,46
-38,5333	-3,7219	2077	257	07/05/2014 06:51	Rua Castro e Silva	42	22,05
-38,5703	-3,7077	2080	2013	07/05/2014 06:58	Av, Presidente Castelo Branco 2	321	22,58
-38,5425	-3,7153	2080	474	07/05/2014 07:13	Av, Presidente Castelo Branco 4	120	14,22
-38,5256	-3,7411	2089	1474	07/05/2014 07:14	Av, Aguanambi 2	1247	4,26
-38,5517	-3,7669	2126	1187	07/05/2014 07:10	Av, Professor gomes de Matos	216	19,79
-38,4849	-3,7256	2214	1811	07/05/2014 07:14	Av, da Abolição 2	402	16,21
-38,5169	-3,7391	2261	1674	07/05/2014 07:03	Av, Padre Valdevino	308	19,56
-38,5327	-3,7220	2290	403	07/05/2014 06:51	Rua Castro e Silva	335	4,33
-38,5055	-3,7245	2311	311	07/05/2014 06:56	Av, Mosenhor Tabosa	30	37,38

Fonte: O autor. Dados: PMF

Tabela 7 - Amostra de indicadores de velocidade média calculados para o grupo de tratamento em maio de 2014

Longitude	Latitude	Id	Dist (m)	Data-Hora	Logradouro	diff (s)	Vm (km/h)
-38,5386	-3,7385	2045	287	14/05/2014 07:01	Av, Carapinima	30	34,48
-38,5369	-3,7353	2056	701	14/05/2014 06:54	Avenida do Imperador	225	11,22
-38,5507	-3,7515	2068	303	14/05/2014 07:11	Av, José Bastos 1	72	15,14
-38,4942	-3,7377	2085	958	14/05/2014 07:02	Avenida Santos Dumont	181	19,06
-38,4977	-3,7470	2087	955	14/05/2014 06:50	Av, Antônio Sales	284	12,10
-38,5001	-3,7462	2154	1.075	14/05/2014 06:52	Av, Antônio Sales	210	18,42
-38,5537	-3,7532	2158	387	14/05/2014 07:14	Av, José Bastos 2	120	11,61
-38,4996	-3,7464	2162	1.488	14/05/2014 07:13	Av, Antônio Sales	504	10,63
-38,4851	-3,7363	2185	291	14/05/2014 06:48	Avenida Dom Luiz	30	34,94
-38,5392	-3,7389	2202	253	14/05/2014 06:52	Av, Carapinima	30	30,38
-38,5523	-3,7513	2202	1.176	14/05/2014 06:58	Av, José Bastos 2	120	35,28
-38,5412	-3,7332	2248	1.935	14/05/2014 07:05	Av, Domingos Olímpio 2	458	15,21

Fonte: O autor. Dados: PMF

Tabela 8 - Amostra de indicadores de velocidade média calculados para o grupo de controle em março de 2015

Longitude	Latitude	Id	Dist (m)	Data-Hora	Logradouro	diff (s)	Vm (km/h)
-38,5522	-3,76749	2037	1292	05/03/2015 06:52	Av, Professor gomes de Matos	283	16,44
-38,5601	-3,77557	2037	400	05/03/2015 06:57	Av, Germano Frank	51	28,20
-38,5434	-3,7586	2040	611	05/03/2015 06:47	Av, Professor gomes de Matos	140	15,71
-38,5236	-3,75298	2044	1665	05/03/2015 06:57	Av, Aguanambi 2	348	17,23
-38,5859	-3,70079	2046	1301	05/03/2015 06:48	Av, Presidente Castelo Branco 2	221	21,20
-38,5249	-3,74017	2048	1592	05/03/2015 07:01	Av, Aguanambi 2	366	15,66
-38,5336	-3,72187	2073	260	05/03/2015 06:45	Rua Castro e Silva	25	37,40
-38,548	-3,76257	2076	560	05/03/2015 07:10	Av, Professor gomes de Matos	90	22,39
-38,5703	-3,70779	2077	2028	05/03/2015 06:56	Av, Presidente Castelo Branco 2	310	23,56
-38,5434	-3,715	2077	406	05/03/2015 07:11	Av, Presidente Castelo Branco 4	59	24,77
-38,5341	-3,72169	2079	239	05/03/2015 07:14	Rua Castro e Silva	30	28,69
-38,5774	-3,70494	2080	1142	05/03/2015 06:56	Av, Presidente Castelo Branco 2	158	26,02
-38,5514	-3,76651	2082	1300	05/03/2015 07:03	Av, Professor gomes de Matos	266	17,60
-38,5235	-3,75191	2095	1515	05/03/2015 06:49	Av, Aguanambi 2	198	27,54

Fonte: O autor. Dados: PMF

Tabela 9 - Amostra de indicadores de velocidade média calculados para o grupo de tratamento em março de 2015

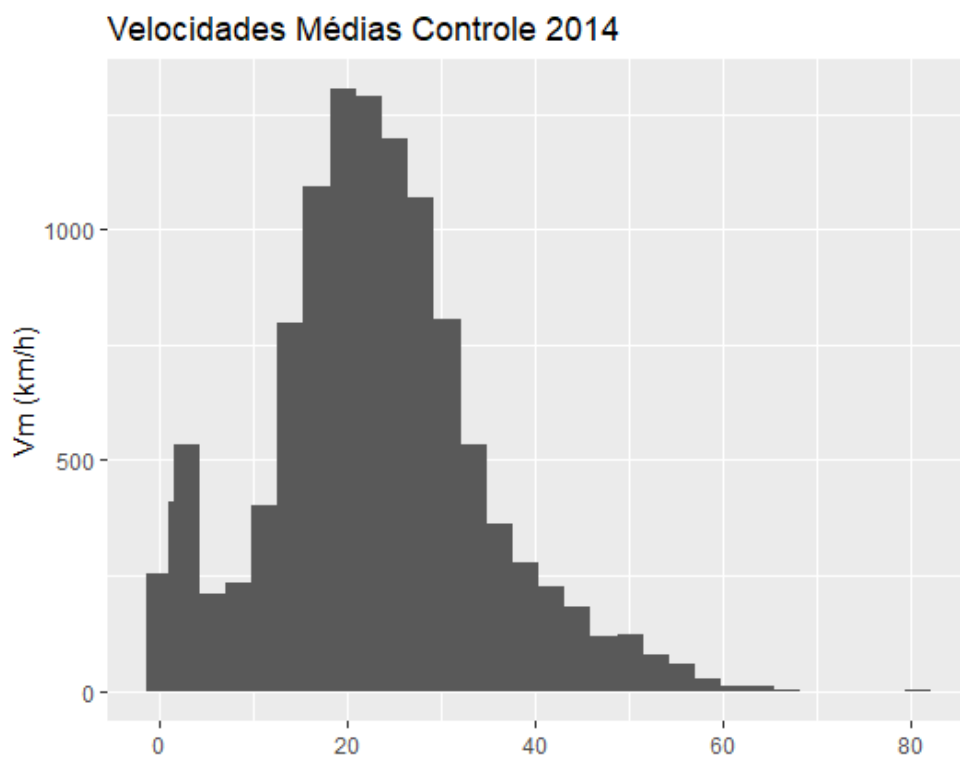
Longitude	Latitude	Id	Dist (m)	Data-Hora	Logradouro	diff (s)	Vm (km/h)
-38,5271	- 3,73782	2022	295	07/03/2015 07:10	Av, Domingos Olímpio 2	72	14,77
-38,5023	- 3,74551	2100	823	07/03/2015 07:06	Av, Antônio Sales	98	30,24
-38,5033	- 3,74524	2103	706	07/03/2015 07:02	Av, Antônio Sales	126	20,17
-38,5367	- 3,73451	2112	296	07/03/2015 07:02	Avenida do Imperador	82	13,00
-38,5281	- 3,73745	2148	429	07/03/2015 07:07	Av, Domingos Olímpio 2	166	9,30
-38,5364	- 3,73376	2155	215	07/03/2015 06:59	Avenida do Imperador	60	12,91
-38,5268	- 3,73805	2155	601	07/03/2015 07:04	Av, Domingos Olímpio 2	101	21,42
-38,5078	- 3,74386	2155	1713	07/03/2015 07:11	Av, Antônio Sales	293	21,04
-38,5024	-3,7455	2156	812	07/03/2015 07:02	Av, Antônio Sales	158	18,50
-38,4995	- 3,74632	2187	1155	07/03/2015 07:04	Av, Antônio Sales	341	12,20
-38,5365	- 3,73406	2194	252	07/03/2015 07:11	Avenida do Imperador	90	10,07
-38,5519	- 3,75136	2216	1301	07/03/2015 06:47	Av, José Bastos 2	102	45,90
-38,5421	- 3,73288	2220	392	07/03/2015 06:48	Av, Domingos Olímpio 2	53	26,63
-38,5157	- 3,74141	2233	291	07/03/2015 07:10	Av, Antônio Sales	29	36,16
-38,4994	- 3,73262	2245	512	07/03/2015 06:46	Avenida Dom Luiz	100	18,43
-38,5575	- 3,75675	2247	2103	07/03/2015 07:00	Av, José Bastos 2	172	44,01
-38,5704	-3,7696	2247	1756	07/03/2015 07:04	Av, Américo Barreiro 2	242	26,13
-38,5386	- 3,73422	2250	1620	07/03/2015 06:58	Av, Domingos Olímpio 2	285	20,47
-38,5027	- 3,74531	2261	2140	07/03/2015 06:59	Av, Antônio Sales	345	22,33
-38,4918	- 3,73847	2263	1121	07/03/2015 06:55	Avenida Santos Dumont	161	25,07
-38,5495	- 3,79257	2347	3363	07/03/2015 06:59	Av, Bernado Manuel 2	349	34,69

Fonte: O autor. Dados: PMF

6.3 Análise Descritiva da Velocidade Média nos Grupos

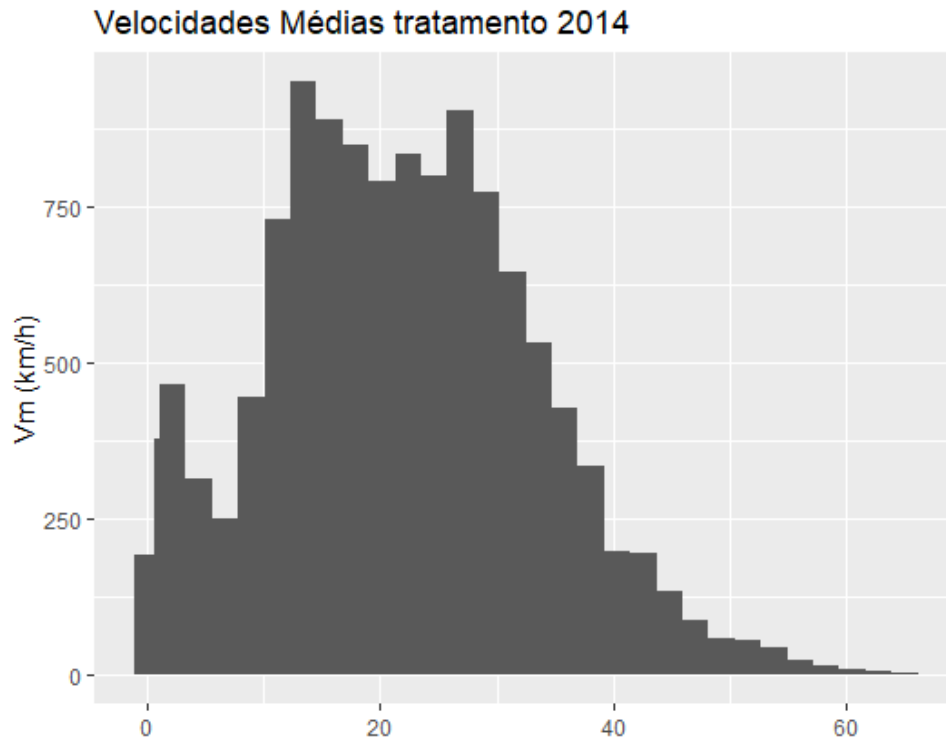
De posse dos indicadores de velocidade média, foi feita uma análise descritiva desse indicador para cada grupo. A distribuição das velocidades médias é mostrada na imagem 28 para cada grupo no formato de diagrama de caixa. Além disso, gerou-se o histograma para a velocidade média em cada grupo para visualizar a forma da distribuição dos dados. Os histogramas são mostrados nas figuras 24, 25, 26 e 27.

Figura 24 - Histograma para a velocidade média do grupo de controle em maio de 2014



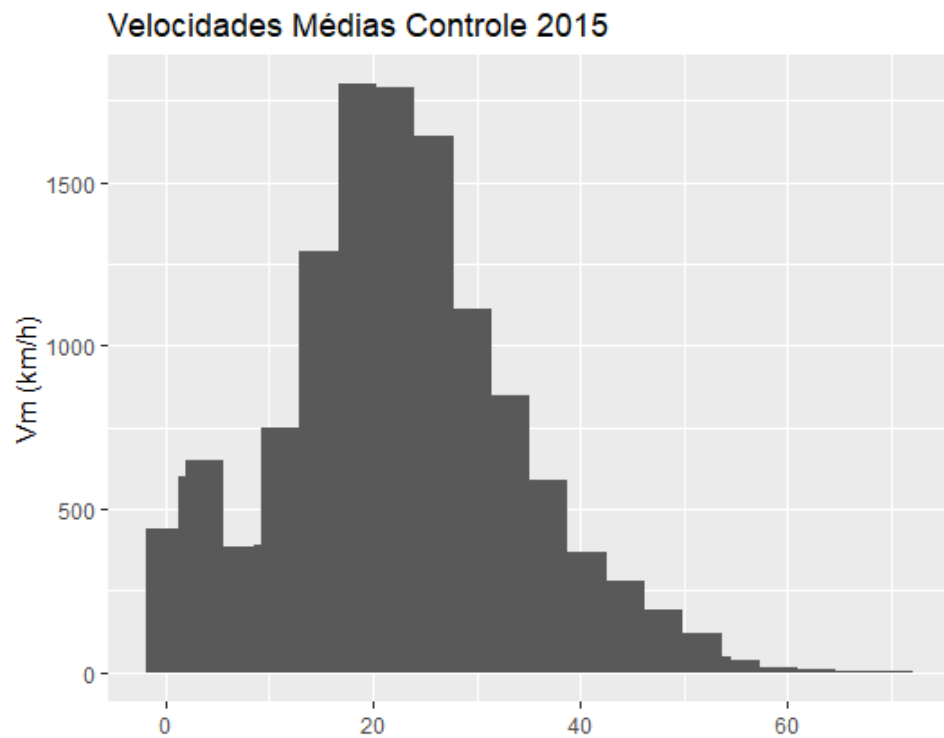
Fonte: O autor. Dados: PMF

Figura 25 - Histograma para a velocidade média do grupo de tratamento em maio de 2014



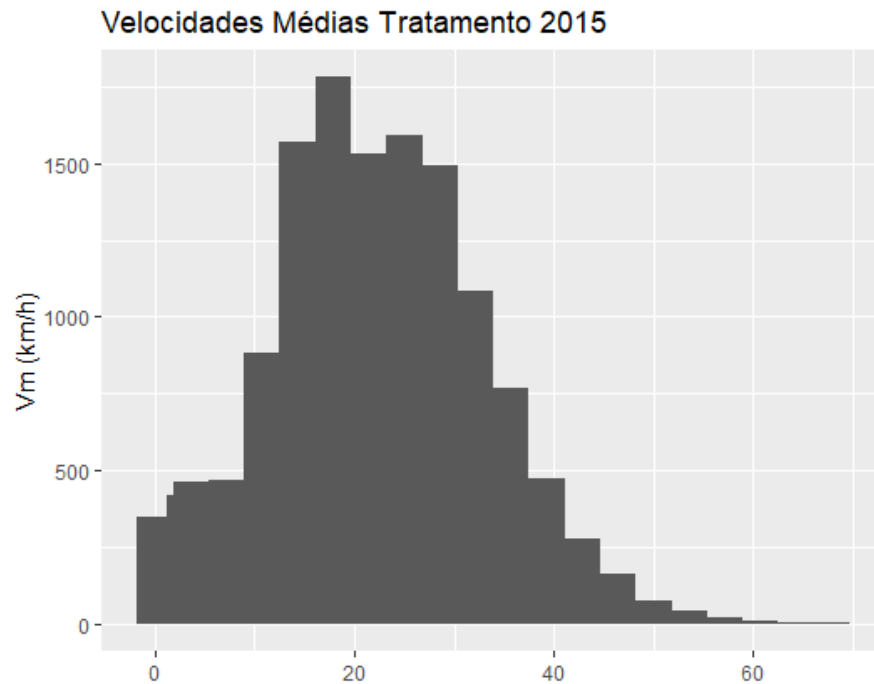
Fonte: O autor. Dados: PMF

Figura 26 - Histograma para a velocidade média do grupo de controle em março de 2015



Fonte: O autor. Dados: PMF

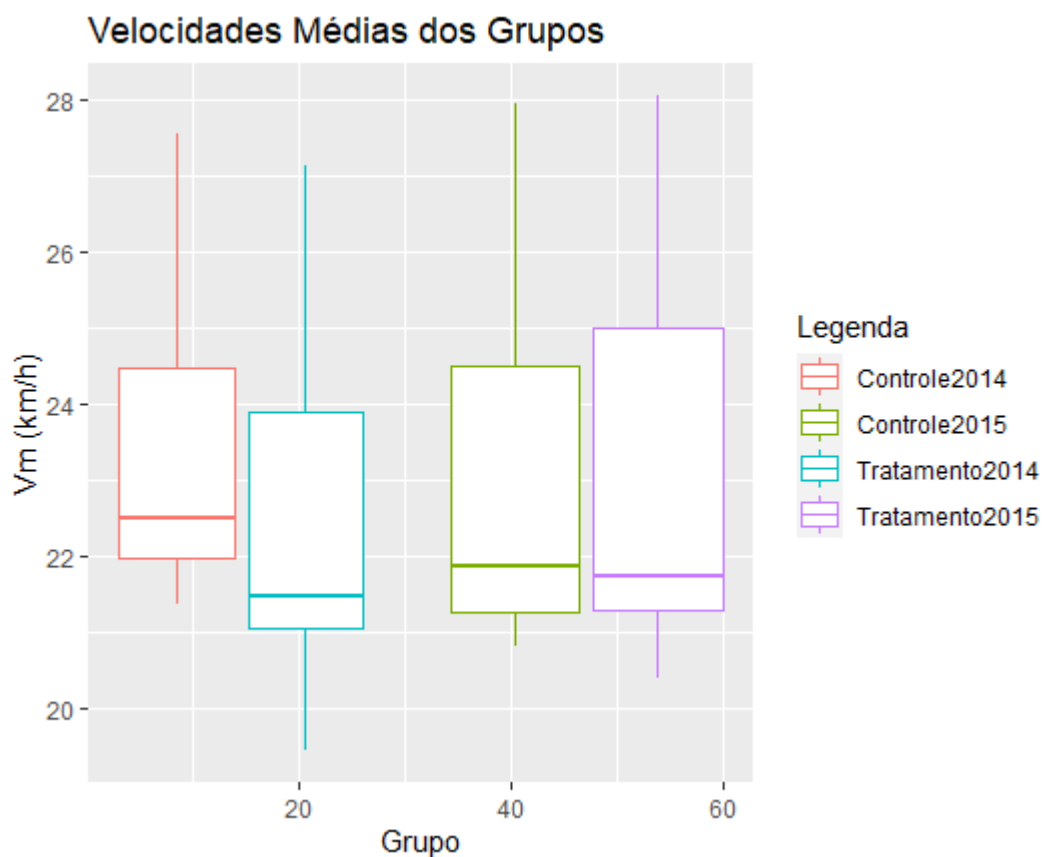
Figura 27 - Histograma para a velocidade média do grupo de tratamento em março de 2015



Fonte: O autor. Dados: PMF

Pelos histogramas é possível observar que a forma é similar para todos os grupos, se assemelhando muito à uma distribuição normal, embora em todos os casos assimétrica à direita (valores mais baixos de velocidade são mais comuns). Além disso, também é possível observar que nos grupos de controle, e no grupo de tratamento antes de receber as faixas exclusivas, há uma concentração de valores de velocidade média em torno de 5 km/h, o que não acontece na distribuição das velocidades médias para o grupo de controle após receber as faixas exclusivas. Isso pode ser um indicativo de que as faixas exclusivas reduzem a ocorrência de velocidades mais baixas.

Passo seguinte, foi calculado as medidas de tendência central e de dispersão para os quatro grupos e dados, mostrado na tabela 10.

Figura 28 - *Box plot* das velocidades dos dois grupos para maio de 2014 e março de 2015

Fonte: o autor. Dados: PMF.

Tabela 10 - Estatísticas de tendência central e dispersão para os grupos de controle e tratamento em maio de 2014 e março de 2015.

Estatística	05/14 Control	05/14 Trat	03/15 Control	03/15 Trat
Min	0,49	0,44	0,45	0,44
1° Quartil	16,39	13,80	15,55	15,11
Mediana	22,59	21,53	22,02	22,01
Média	23,08	22,00	22,56	22,46
Terceiro Quartil	29,27	29,53	29,30	29,48
Max	81,20	65,44	70,73	68,38
Desvio Padrão	11,31	11,34	11,32	10,66
Variância	127,91	128,51	128,14	113,61
Tamanho da Amostra (n)	11206	11915	12314	13043

Fonte: o autor.

6.4 Impacto das Faixas Exclusivas nos Indicadores

Por fim, foi aplicado o método da diferença nas diferenças explicitado na seção 5.4.

6.4.1 Impacto na Velocidade Média

Obteve-se uma diferença de $V_m = 1,78$ km/h entre os grupos, correspondente à aproximadamente 7,9 % da velocidade média no grupo tratado. Foi preciso verificar essa medida quanto à significância estatística. Para isso, foi feito um teste t de Welch para variáveis não pareadas, com hipótese alternativa de a diferença encontrada é superior a zero. O teste teve valor de confiança muito acima do estipulado de 95%, com valor $p = 6,46 \times 10^{-6}$. Portanto, a diferença encontrada, apesar de pequena, é significativa.

O teste t foi feito no *RStudio*, e os valores retornados são mostrados na figura 29.

Figura 29 - Teste t para a dupla diferença de velocidade média entre o grupo tratado e o grupo de controle

```
welch Two Sample t-test
data:  vm1$D1 and vm0$D0
t = 4.3905, df = 756.53, p-value = 6.46e-06
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 1.110854      Inf
sample estimates:
mean of x  mean of y
1.3386870 -0.4389648
```

Fonte: O autor, com uso do Rstudio.

Embora o impacto medido pela diferença nas diferenças seja pequeno, ele confirma que a diferença é positiva, ou seja, as faixas exclusivas têm um impacto positivo na velocidade operacional dos ônibus. Contudo, o resultado quanto à magnitude desse resultado pode estar incerto devido a variáveis não consideradas no modelo, bem como pode haver variáveis mediadoras que reduzem o efeito total das faixas exclusivas. Isto é conhecido como efeito direto, efeito indireto e efeito total na literatura. A separação dos efeitos diretos e indiretos não fez parte do escopo deste trabalho.

6.4.2 Impacto na Variação da Velocidade Média

Os indicadores de velocidade média, que estão agrupados por veículo e por dia, foram agrupados por corredor por dia. Dessa forma, foi somado a variância das linhas que passaram pelo corredor em questão. Porém, como supradito, espera-se que essa variância intrínseca às linhas seja cancelada ao fazer a diferença nas diferenças.

Com as variâncias calculadas para cada corredor, procedeu-se conforme o método em 5.3, calculando o valor esperado das variâncias para os dois grupos. Em seguida, calculou-se as diferenças. O resultado foi uma redução de -2,6422 (km/h) para o grupo tratado. Novamente, procedeu-se a verificação da significância desse resultado. Para isso, foi feito o teste F no Rstudio, com hipótese alternativa de que a razão entre as variâncias é inferior à 1, ou seja, a variância da velocidade média do grupo tratado é inferior à do grupo não tratado. Novamente, houve significância estatística no resultado, bem acima de 95% e valor p de $6,73 \times 10^{-5}$.

Os resultados do teste F feito no Rstudio são mostrados na imagem 30.

Figura 30 - Teste F para variância das velocidades nos dois grupos

```
F test to compare two variances
data:  vmv1$D1 and vmv0$D0
F = 0.66568, num df = 324, denom df = 408, p-value = 6.734e-05
alternative hypothesis: true ratio of variances is less than 1
95 percent confidence interval:
 0.0000000 0.7924598
sample estimates:
ratio of variances
 0.6656826
```

Fonte: o autor, com uso do RStudio.

Além do teste de significância, é importante notar que o intervalo de confiança para a razão das variâncias é bem distante do valor 1 (variâncias iguais), com valor médio de razão de 0,67, favorecendo à variância do grupo tratado.

Novamente, o ganho para o grupo tratado foi pequeno, o que pode sugerir que pode haver outras variáveis importantes que não foram consideradas no modelo.

6.5 Impacto das faixas exclusivas utilizando DiD com controle para o volume

Utilizando o controle para volume na regressão, obteve-se redução no impacto das faixas exclusivas e o impacto foi estatisticamente significativo. Observa-se ainda que a influência do volume na velocidade média foi praticamente inexistente, fato contrário ao que se esperava. Isso pode indicar que a escolha dos grupos não carrega diferenças no volume ou que a velocidade média no grupo de tratamento não sofre influência do volume do corredor.

O resultado da análise com controle para volume é mostrado na figura 31.

Figura 31 - Resultado do Modelo com Controle para Volume no Corredor

term	estimate	std.error	statistic	p.value
<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1 tratamento	1.11	0.236	4.70	0.00000264
2 volume	-0.00000197	0.0000461	-0.0427	0.966

Fonte: O autor.

7 CONCLUSÕES E RECOMENDAÇÕES

Conforme o primeiro objetivo específico, foi possível levantar indicadores de desempenho para faixas com e sem faixa exclusiva. Foi possível estimar a velocidade média dos veículos nos trechos escolhidos e sua variância. Quanto ao headway, foi proposto um método para calculá-lo, contudo não foi possível implementá-lo devido ao grande custo somente para calcular o indicador de velocidade média. Além disso, como não foi possível pôr em prática o método adicional de integração das bases de dados de GPs e Bilhetagem, utilizando a base de GTFS. Dessa forma, apesar das dificuldades encontradas para o processamento dos dados massivos, foi possível calcular o indicador de velocidade média e utilizá-lo para avaliar o impacto devido às faixas exclusivas em Fortaleza, cumprindo o objetivo principal deste trabalho.

Nesse contexto, no decorrer deste trabalho foi constatado que há evidências que corroboram um efeito positivo na velocidade média dos veículos bem como sua redução, embora os efeitos medidos tenham sido pequenos (aproximadamente 1,8 km/h para a velocidade média, ou seja, 7,9% e redução de 2,6 km²/h² na variância), mas significativos. Dessa forma, constatou-se também que a magnitude dos ganhos nos indicadores verificados estão muito aquém dos divulgados pela prefeitura de Fortaleza, a qual chegou a anunciar ganhos em velocidade acima de 100% em alguns trechos, se utilizando apenas de comparação antes e depois. Uma explicação para isso pode ser a de que as faixas exclusivas têm um impacto grande no curtíssimo prazo (logo após serem implantadas) e com o tempo o sistema por si só trata de entrar em equilíbrio, se utilizando de algum mecanismo, como a redistribuição da demanda pelo transporte público, que não foi avaliada neste trabalho.

Também não foi avaliado neste trabalho a influência de algumas variáveis, devido à dificuldade de coletá-las, especialmente em um período de pandemia, que o autor acredita que possuam grande influência nos indicadores, como a distribuição de semáforos, uso do solo, distância entre paradas e demanda do transporte público. Neste contexto, uma recomendação para futuros trabalhos futuros é de incluir mais variáveis de controle no modelo, ou ainda verificar o impacto das faixas exclusivas no tráfego geral, e vice-versa.

Uma outra conclusão é de que dificilmente um modelo de inferência causal que não permita estimar o impacto direto e indireto seja capaz de dar uma estimativa confiável da magnitude dos impactos constatados. Ressalta-se ainda que nenhum estudo do tipo deve ser considerado como conclusivo, cabendo esse papel ao conjunto de trabalhos que concluírem a mesma tendência de impacto, por métodos diferentes. Isso dito, constatou-se que o ganho na

velocidade operacional para o grupo tratado foi bem similar ao encontrado no trabalho de Arbex e Da Cunha (2016) para as faixas exclusivas da cidade de São Paulo (9,13%), os quais não utilizaram um método causal explícito, mas fizeram uma comparação similar.

Conclui-se ainda, em linha com o segundo e terceiro objetivo específico, que foi possível determinar um método de inferência causal adequado aos dados massivos de transporte do SIT-FOR, bem como foi possível estimar um impacto devido às faixas exclusivas com um maior grau de confiança relação a uma simples comparação antes e depois, como geralmente feito pelas empresas de transporte.

Por fim, deve-se salientar que o fato de os dados de GPS e bilhetagem não serem compatíveis sem que antes seja feito todo um método de consolidação traz muito prejuízos às análises de transporte, tanto no tempo computacional necessário para somente consolidá-los e nos registros que acabam por ser descartados por não serem possíveis de consolidar. Nesse sentido, vale citar o trabalho de Arbex e Da Cunha (2016), no qual foi utilizado dados de GPS nativamente consolidados, ou seja, gerados já com as informações da bilhetagem necessárias para retirar uma quantidade de informação maior dos dados. Dessa forma, conclui-se que a compatibilidade entre essas bases de dados é possível quando os sistemas são implantados pensando não somente em reduzir custos ou aumentar segurança, mas também aproveitar esses dispositivos para gerar dados secundários com grande valor para produzir análises que subsidiem a tomada de decisão.

REFERÊNCIAS

- ALRT (ORG). Agência de Transportes do Cazaquistão. **Exclusive Bus Lanes**. Disponível em: <<http://alrt.kz/en/project/18>>. Acesso: 25 fev. 2020.
- ARBEX, R. O.; Da Cunha, C. B. **Avaliação das mudanças nas velocidades das linhas de ônibus da cidade de São Paulo após a implantação de faixas exclusivas através da análise de dados de GPS**. *Transportes*, v. 24, n. 4, p. 21, 2016. ISSN 2237-1346. Disponível em: <<https://www.revistatransportes.org.br/anpet/article/view/1008>>.
- BARAT, JOSEF; BATISTA, M. S. N. **Transporte público e programas habitacionais**. 1973. Disponível em: <<http://repositorio.ipea.gov.br/handle/11058/6582>>.
- BRAGA, C. **Big Data de Transporte Público na Análise da Variabilidade de Indicadores da Acessibilidade às Oportunidades de Trabalho e Educação**. 105 f. Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Transportes.
- BRINCKERHOFF, P. **Transit Capacity and Quality of Service Manual**, Third Edition.[S.l.], 2013. Disponível em: <<https://www.nap.edu/catalog/24766>>.
- BURINSKIENĖ, M.; GUSAROVIEŅĖ, M.; GABRULEVIČIŪTĖ-SKEBIENĖ, K. **The impact of public transport lanes on the operating speed of buses**. 9th International Conference on Environmental Engineering, ICEE 2014, 2014. n. May, p. 1–6.
- CAMPBELL, K. B.; BRAKEWOOD, C. **Sharing riders: How bikesharing impacts bus ridership in New York City**. *Transportation Research Part A: Policy and Practice*, 2017.
- Caracterização da Problemática do Sistema de Transporte Público**. Tese (Dissertação de Mestrado) — Universidade Federal do Ceará, 2017.
- CARTNER. **Information assets that are characterized by massive, diversified, and high growth rates that require new processing models for greater decision-making, insight and process optimization capabilities**. 2001.

CORTÉS, C. E.; GIBSON, J.; GSCHWENDER, A.; MUNIZAGA, M.; ZÚÑIGA, M. **Commercial bus speed diagnosis based on GPS-monitored data**. Transportation Research Part C: Emerging Technologies, v. 19, n. 4, p. 695–707, 2011. ISSN 0968090X.

DENG, T.; NELSON, J. D. **Recent developments in bus rapid transit: A review of the literature**. Transport Reviews, 2011. v. 31, n. 1, p. 69–96.

FORTALEZA. PREFEITURA MUNICIPAL DE FORTALEZA. (org.). **Faixas Exclusivas de Ônibus**. 2018. Disponível em: <https://mobilidade.fortaleza.ce.gov.br/menu-programas/faixa-exclusivas-de-%C3%B4nibus.html>. Acesso em: 04 fev. 2021.

GAO, S.; LI, L.; GOODCHILD, M. F. **A Scalable Geoprocessing Workflow for Big Geo-Data Analysis and Optimized Geospatial Feature Conflation based on Hadoop**. 2013.

GELMAN E HILL. **Data Analysis Using Regression and Multilevel/Hierarchical Models**. Disponível em: https://books.google.com.br/books?hl=fr&lr=&id=c9xLKzZWoz4C&oi=fnd&pg=PR17&dq=Gelman+%26+Hill,+2007&ots=bcNbN3Prsb&sig=PJ9S-JTS3HBXmTs0SM5RLmtvZSc&redir_esc=y#v=onepage&q=Gelman%20%26%20Hill%20C%202007&f=false 2006. Acesso em 20 mai. 2020.

GERTLER, Mark, Nobuhiro Kiyotaki, **Chapter 11 - Financial Intermediation and Credit Policy in Business Cycle Analysis**. Benjamin M. Friedman, Michael Woodford, Handbook of Monetary Economics. Elsevier, Volume 3, 2010, Pages 547-599, ISSN 1573-4498, ISBN 9780444534705, <https://doi.org/10.1016/B978-0-444-53238-1.00011-9>. Disponível em: <https://www.sciencedirect.com/science/article/pii/B9780444532381000119>>

IPEA. **Política Nacional de Mobilidade Urbana, Lei 12.587/12**. Ministério das Cidades, 2013. p. 1–37. Disponível em: http://www.secid.ma.gov.br/files/2015/03/cartilha_lei_12587-1.pdf>.

IPEA. **Transporte Urbano e Inclusão Social: Elementos para Políticas Públicas. Ação para a Expansão do Metro-Ferroviário nas Regiões Metropolitanas.** *Instituto de pesquisa econômica aplicada*, p. 1–26, 2003.

KHANDKER, Shahidur & Koolwal, Gayatri & Samad, Hussain & Bank, World. (2010). **Handbook On Impact Evaluation: Quantitative Methods And Practices.** [http://lst-iiiep.iiep-unesco.org/cgi-bin/wwwi32.exe/\[in=epidoc1.in\]/?t2000=028739/\(100\).10.1596/978-0-8213-8028-4](http://lst-iiiep.iiep-unesco.org/cgi-bin/wwwi32.exe/[in=epidoc1.in]/?t2000=028739/(100).10.1596/978-0-8213-8028-4).

KITCHIN, R. **Big data and human geography: Opportunities, challenges and risks.** *Dialogues in Human Geography*, v. 3, n. 3, p. 262–267, 2013. ISSN 20438214. *Land Use Policy*, 15 pp. 233-251.

LIU SHUO. **Geoprocessing Optimization in Grids.** 2005. Disponível em: <<http://d-scholarship.pitt.edu/8677/>>.

MORA, R.; REGGIO, I. **Treatment Effect Identification Using Alternative Parallel Assumptions.** *Universidad Carlos III de Madrid Working Papers*, 2013. v. 12–33, n. November, p. 1–28.

NISIDA, V. C. ; VANNUCHI, L. V. B., *et al.* **A inserção urbana dos empreendimentos do programa Minha Casa Minha Vida na escala local: uma análise do entorno de sete conjuntos habitacionais.** *Xvi Enanpur*, p. 24, 2014.

OLIVEIRA, João Lucas Albuquerque. **Análise da Evolução da Demanda no Transporte Coletivo por Ônibus em Fortaleza Utilizando Big Data.** 2019. 70 f. TCC (Graduação) - Curso de Engenharia Civil, Centro de Tecnologia, Universidade Federal do Ceará, Fortaleza, 2019.

OLIVEIRA-NETO, F. M.; LOUREIRO, C. F. G.; HAN, L. D. **Active and passive bus priority strategies in mixed traffic arterials controlled by SCOOT adaptive signal system: Assessment of performance in Fortaleza, Brazil.** *Transportation Research Record*, 2009. n. 2128, p. 58–65.

PANGILINAN, C., Wilson, N. and Moore, A. (2008) '**Bus Supervision Deployment Strategies and Use of Real-Time Automatic Vehicle Location for Improved Bus Service Reliability**', *Transportation Research Record*, 2063(1), pp. 28–33. doi: 10.3141/2063-04.

PELLETIER, M. P.; TRÉPANIER, M.; MORENCY, C. **Smart card data use in public transit: A literature review**. *Transportation Research Part C: Emerging Technologies*, Elsevier Ltd, v. 19, n. 4, p. 557–568, 2011. ISSN 0968090X.

PERCOCO, M. **Environmental Effects of the London Congestion Charge: a Regression Discontinuity Approach**. 2017. Disponível em: <https://transport.epfl.ch/heart/2015/abstracts/hEART_2015_submission_114.pdf>.

PORTAL ARCHIDAILY (ORG). **BRT de Curitiba é eleito um dos projetos mais influentes do mundo nos últimos 50 anos**. Disponível em: <<https://www.archdaily.com.br/br/931784/brt-de-curitiba-e-eleito-um-dos-projetos-mais-influentes-do-mundo-nos-ultimos-50-anos>> . Acesso em 25 fev. 2020.

PORTAL DIÁRIO DO TRANSPORTE. **Bilhetagem Eletrônica: Cadastro em Fortaleza e Santos só a dinheiro**. Disponível em: <https://diariodotransporte.com.br/2013/04/21/bilhetagem-eletronica-cadastro-em-fortaleza-e-santos-so-a-dinheiro/>. Acesso em 25 mar. 2020.

PORTAL FORTALBUS (ORG). **Fortaleza Implanta Vias de Sentido Único**. Disponível em: < fortalbus.com/2014/05/fortaleza-implantacao-de-sentido-unico.html>. Acesso em 25 fev. 2020.

PORTAL MOBILIZE (ORG). **Retomadas obras de mobilidade urbana em Fortaleza**. Disponível em: <<https://www.mobilize.org.br/noticias/2647/obras-de-mobilidade-para-a-copa-em-fortaleza.html>>. Acesso em 25 mar. 2020.

PORTAL O POVO (ORG). **Motoristas desrespeitam faixas exclusivas de ônibus.** Disponível em:

<<https://www20.opovo.com.br/app/opovo/fortaleza/2012/08/08/noticiasjornalfortaleza,2894802/motoristas-desrespeitam-faixas-exclusivas-para-onibus.shtml>> Acesso em 25 mar 2020.

QUIROGA, C. A.; BULLOCK, D. **Travel time studies with global positioning and geographic information systems: an integrated methodology.** Transportation Research Part C: Emerging Technologies, v. 6C, n. 1-2, p. 101–127, 1998. ISSN 0968090X.

RABAY, P. d. F. B. **Uso de Dados Secundários de Rastreamento da Frota na Etapa de Caracterização da Problemática do Sistema de Transporte Público.** Tese (Dissertação de Mestrado) — Universidade Federal do Ceará, 2017.

RABINOVITCHE LEITMAN. (1996) **Urban planning in Curitiba.** Scientific American, 274(3):46–53.

RUBIN, D. B. (1974). **Estimating causal effects of treatments in randomized and nonrandomized studies.** Journal of Educational Psychology, 66(5), 688–701. <https://doi.org/10.1037/h0037350>.

SALAS, C. **Evaluating Public Policies with High Frequency Data: Evidence for Driving Restrictions in Mexico City Revisited.** Economia IEUC, 2010.

SCHIFF, A.; WRIGHT, L.; DENNE, T. **Ex-post evaluation of transport interventions using causal inference methods.** Wellington: NZ Transport Agency, 2017.

SMITH, H and Raemaekers, J. (1998). **Land use pattern and transport in Curitiba.**

SOUZA, E., Maria, R. De, Bernadete, C., Cruz, M., & Richter, M. (2014). **O Uso De Geotecnologias Em Sistemas De Transporte E Organização Urbana No Brasil.** Mercator : Revista de Geografia Da UFC, 13(1), 143–152. <https://doi.org/10.4215/rm.v13i1.918>

STORCK, Gerson Luís. **Pesquisas em ônibus urbano.** 1988. 86 f. TCC (Graduação) - Curso de Estatística, Universidade Federal do Rio Grande do Sul, Porto Alegre, 1988.

TIRACHINI, A. **The economics and engineering of bus stops : Spacing , design and congestion.** Transportation Research Part A, 2014. v. 59, p. 37–57. Disponível em: <<http://dx.doi.org/10.1016/j.tra.2013.10.010>>.

TOWNES, M. S.; BARNES, L. E. E.; BLAIR, G. L.; HUNTER-ZAWORSKI, K.; MONROE, D. O. N. S. **AVL Systems for Bus Transit: A Synthesis of Transit Practice.** [S.l.]: Transit Cooperative Research Program, 1997. ISBN 0309060192.

UKPTA - UK PUBLIC TRANSPORT AUTHORITY. **2010 to 2015 government policy: local transport.** Disponível em: <<https://www.gov.uk/government/publications/2010-to-2015-government-policy-local-transport/2010-to-2015-government-policy-local-transport>>. Acesso em: 25 fev. 2020.

US DEPT. OF DEFENSE (ORG). **What on Earth Is The Global Positioning System.** Disponível em: <<https://www.defense.gov/Explore/Features/story/Article/1674004/what-on-earth-is-the-global-positioning-system/>>. Acesso em 13 abr. 2021.

VINSENTIN, J. M.; LUCCHESI, S. T.; SILVEIRA, I. M. **Efeitos da inserção urbana e periferização no transporte coletivo urbano: estudo de caso em um conjunto habitacional do programa minha casa minha vida.** 33º Congresso de Pesquisa e Ensino em Transporte da ANPET, 2019. v. 53, n. 9, p. 1689–1699.

VUCHIC, V. R.; ENGINEERING, S. **Urban public transportation systems.** [s.d.]. p. 1–20. 2007.

WARD, J. S.; BARKER, A. **Undefined By Data: A Survey of Big Data Definitions.** 2013. Disponível em: <<http://arxiv.org/abs/1309.5821>>.

ZHAO, J.; RAHBEE, A.; WILSON, N. H. **Estimating a rail passenger trip origin-destination matrix using automatic data collection systems.** Computer-Aided Civil and Infrastructure Engineering, v. 22, n. 5, p. 376–387, 2007. ISSN 10939687.