



VARIAÇÃO DO ÍNDICE DE MOBILIDADE SEMANAL POR MEIO DE DADOS
DO MOZILLA LOCATION SERVICE

Frank Lennox Max Jonnes de Oliveira Queiroz

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia de Transportes, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia de Transportes.

Orientador: Paulo Cezar Martins Ribeiro

Rio de Janeiro

Março de 2018

VARIAÇÃO DO ÍNDICE DE MOBILIDADE SEMANAL POR MEIO DE DADOS
DO MOZILLA LOCATION SERVICE

Frank Lennox Max Jonnes de Oliveira Queiroz

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO
LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA
(COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE
DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE
EM CIÊNCIAS EM ENGENHARIA DE TRANSPORTES.

Examinada por:

Prof. Paulo Cezar Martins Ribeiro, Ph.D.

Prof. Felipe Maia Galvão França, Ph.D.

Prof. Diego Moreira de Araujo Carvalho, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

MARÇO DE 2018

Queiroz, Frank Lennox Max Jonnes de Oliveira

Varição do Índice de Mobilidade Semanal por Meio de Dados do *Mozilla Location Service* – Rio de Janeiro: UFRJ/COPPE, 2018.

XIII, 95 p.: il.; 29,7 cm.

Orientador: Paulo Cezar Martins Ribeiro

Dissertação (mestrado) – UFRJ/ COPPE/ Programa de Engenharia de Transportes, 2018.

Referências Bibliográficas: p. 86-91.

1. Índice de Mobilidade. 2. *Mozilla Location Service*.
3. Varição. I. Ribeiro, Paulo Cezar Martins. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Transportes. III. Título.

AGRADECIMENTOS

Me lembro do meu primeiro ano como aluno da Engenharia Civil pela Universidade Federal de Mato Grosso. Em um trabalho de química conheci uma instituição chamada COPPE que se situava no Rio de Janeiro. A partir daí sempre a considerei como um centro de excelência, ainda que não soubesse que o mundo inteiro achava o mesmo. Muitas coisas passaram na minha cabeça naquele curto espaço em que conheci a COPPE pela internet, entretanto, a considerava quase que inatingível. Agora estou aqui. Dito isso, dá pra imaginar o tamanho do sonho que realizo ao chegar a este momento tendo estudado no maior centro de engenharia da América Latina.

Não chegaria até aqui sem a ajuda de inúmeras pessoas, entre as quais, os mais importantes foram meus pais, Gilberto e Eula. Apoiaram-me nas grandes dificuldades iniciais, em que saí pela primeira vez de Cuiabá já para morar no Rio de Janeiro. Minha irmã, Poliana, mestre em antropologia, teve importância semelhante, visto que sem ela talvez não teria optado pelo mestrado. Foram tempos árduos, mas vocês foram as grandes motivações da busca pela vitória.

Não apenas seres humanos fizeram parte dessa caminhada, mas também meus cachorros, devo agradecimentos pelas recepções de Pink e Toddy cada vez que voltava pra Cuiabá, bem como as brincadeiras diárias em que eu sempre cansava primeiro. Objetos também fizeram parte, um em especial, meu saudoso e afinado violão.

Nestes dois anos fiz ou fortaleci laços de amizades fantásticos. Dênis Guedes, que ajudou a carregar minhas malas de um Frank ainda perdido na Vila Universitária. Wesley Braytiner, personificação do carioca, foi meu companheiro de pesquisa, me deu a grande motivação para seguir determinado e focado. Marina Baltar, moça de grande coração, não mediu esforços para me ajudar dentro e fora da academia. Thales Braga, que já era meu amigo há tempos, me propiciou o primeiro contato com os dados do Mozilla Location Service. Castieli, Daniel e George, amigos da boemia, me deram leveza aos tempos mais complicados, assim como os amigos que morei, Darlan e Roger. Amigos com quem trabalhei, Luiz Afonso, Livia e Rebeca, companheiros de Laboratório ITS. Amigos de sala, com quem tive muitos momentos agradáveis no primeiro ano de estudo.

Da COPPE, devo agradecimentos a Heleninha e a Jane pelos serviços prestados com muito gentileza e alegria. Às meninas do Burguesão, pela simpatia e sorriso no rosto em cada um dos muitos de cafés que lá comprei. Também, é claro, devo

agradecimentos ao corpo docente do Programa de Engenharia de Transportes por facilitar a busca incessante pelo conhecimento. Um agradecimento especial ao meu orientador, Prof. Paulo Cezar, pela paciência e direcionamentos que foram essenciais a conclusão satisfatória deste trabalho.

Por fim, agradeço ao patrocínio do CNPq que possibilitou a elaboração deste trabalho.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

VARIAÇÃO DO ÍNDICE DE MOBILIDADE SEMANAL POR MEIO DE DADOS DO MOZILLA LOCATION SERVICE

Frank Lennox Max Jonnes de Oliveira Queiroz

Março/2018

Orientador: Paulo Cezar Martins Ribeiro

Programa: Engenharia de Transportes

Este estudo desenvolve uma abordagem acerca da variação da mobilidade entre dias úteis, finais de semana e feriados. Para este fim, foi utilizada uma fonte alternativa de dados voltados a transporte, o Mozilla Location Service, serviço que fornece dados locais de seus colaboradores. Para o estudo da variação, um método foi construído no intuito de analisar o comportamento do coletivo, sem focar em padrões individuais. Não se encontrou diferenças significativas no índice de mobilidade entre dias úteis, bem como entre dia útil e sábado, já nas comparações de dia útil com domingo e feriados houve variação significativa. Percebeu-se uma tendência de menor mobilidade (descanso) em feriados na segunda-feira, comparados ao de quinta-feira. Os resultados encontrados aqui corroboram, mas também divergem de vários estudos da literatura.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

VARIATION OF THE WEEKLY MOBILITY INDEX BY MOZILLA LOCATION
SERVICE DATA

Frank Lennox Max Jonnes de Oliveira Queiroz

March/2018

Advisor: Paulo Cezar Martins Ribeiro

Department: Engenharia de Transportes

This study develops an approach to variability of mobility between weekdays, weekends, and holidays. For this purpose, an alternative source of transport-oriented data was used, the Mozilla Location Service, which provides locational data for its collaborators. For the study of variation, a method was constructed in order to analyze the behavior of the collective, without focusing on individual patterns. No significant differences were found in the mobility index between working days, as well as between working days and Saturdays, and in the comparisons of working days with Sunday and holidays there was a significant variation. There was a trend of less mobility (rest) on holidays on Monday, compared to Thursday. The results found here corroborate, but also diverge from several studies in the literature.

SUMÁRIO

1.	INTRODUÇÃO.....	1
1.1.	Contextualização.....	1
1.2.	Problema de pesquisa.....	5
1.3.	Hipóteses.....	5
1.4.	Objetivo geral	6
1.5.	Objetivos específicos	6
1.6.	Justificativa	6
1.7.	Estrutura do trabalho.....	7
2.	REVISÃO DE LITERATURA	8
2.1.	Métodos tradicionais de coleta de dados em transportes.....	8
2.1.1.	Pesquisa documental	10
2.1.2.	Pesquisa observacional	10
2.1.3.	Pesquisa de auto resposta	11
2.1.4.	Pesquisa por telefone	11
2.1.5.	Pesquisa de interceptação	11
2.1.6.	Pesquisa domiciliar por entrevista.....	12
2.2.	Métodos alternativos de coleta de dados em transporte.....	13
2.2.1.	Estudos voltados ao transporte público	14
2.2.2.	Estudos baseados em redes sociais.....	15
2.2.3.	Estudos baseados em telefonia móvel	16
2.3.	Variabilidade em transportes	18
2.4.	Estudos acerca da variabilidade da mobilidade	20
2.4.1.	Variabilidade com dados obtidos por métodos tradicionais.....	20
2.4.2.	Variabilidade com dados obtidos por métodos alternativos.....	25
3.	MOZILLA LOCATION SERVICE	30
3.1.	App Mozilla Stumbler	30
3.2.	Registro e atualização dos dados obtidos pelos colaboradores.....	32
3.3.	Caracterização dos dados fornecidos pelo MLS.....	36
4.	MÉTODO PROPOSTO.....	45
4.1.	Considerações adotadas	45
4.2.	Tratamento dos dados	46
4.2.1.	Obtenção dos dados	46
4.2.2.	Filtragem dos dados.....	47
4.2.3.	Atribuição de uma área às coordenadas geográficas	48
4.2.4.	Agregação dos dados	49
4.2.5.	Contagem dos dados.....	51
4.3.	Análise dos dados	52
5.	ESTUDO DE CASO	54
5.1.	Área de estudo	54
5.1.1.	Área de Planejamento 1	55
5.1.2.	Área de Planejamento 2	56
5.1.3.	Área de Planejamento 3	56
5.1.4.	Área de Planejamento 4	57
5.1.5.	Área de Planejamento 5	57
5.2.	Obtenção dos dados	58
5.3.	Filtragem dos dados para o Rio de Janeiro	60
5.4.	Atribuição de áreas para cada registro	61
5.5.	Agregação dos dados	62
5.6.	Contagem das viagens efetuadas	63

5.7.	Análise estatística dos índice de mobilidade	63
6.	RESULTADOS E DISCUSSÕES	65
6.1.	Análise gráfica preliminar.....	65
6.2.	Variabilidade entre dias úteis.....	69
6.3.	Variabilidade entre dia útil e final de semana.....	72
6.3.1.	Comparação do IM entre dia útil e Sábado	72
6.3.2.	Comparação entre IM de domingo e dia útil	74
6.4.	Variabilidade entre dia útil e feriados	75
6.4.1.	Comparação entre dia útil e 1º de Maio	76
6.4.2.	Comparação entre dia útil e Corpus Christi (15/jun).....	77
6.4.3.	Comparação entre dia útil e 16/jun.....	78
6.4.4.	Comparação entre feriados	79
6.5.	Cenário recomendado para uso dos dados MLS.....	81
7.	CONCLUSÃO E RECOMENDAÇÕES	83

LISTA DE FIGURAS

Figura 1 - Hierarquia de decisão para escolha do método de coleta de dados	9
Figura 2 - Células omnidirecionais.....	17
Figura 3 - Células setorizadas.....	17
Figura 4 - Primeira estimativa de localização da antena	34
Figura 5 - Décima estimativa de localização da antena.....	35
Figura 6 - Centésima estimativa de localização da antena	35
Figura 7 - Triangulação de antenas.....	36
Figura 8 - Trajeto de ida	37
Figura 9 - Trajeto de volta	37
Figura 10 - <i>Screenshot</i> no momento do teste	37
Figura 11 - Pontos coletados pelo colaboradores	38
Figura 12 - Processo de estimação da localização da antena	39
Figura 13 - Plotagem dos pontos fornecidos pelo MLS	41
Figura 14 - Antenas da região que foi realizada a coleta.....	42
Figura 15 - Atribuição de área às coordenadas geográficas	48
Figura 16 - Agregação dos dados MLS	50
Figura 17 - Fluxograma do método	53
Figura 18 - Áreas de Planejamento do Rio de Janeiro	55
Figura 19 - Coordenadas do Rio de Janeiro	60
Figura 20 - Aproximação das APs em retângulos	61
Figura 21 – Método recomendado para uso de dados MLS	82

LISTA DE TABELAS

Tabela 1 - Distinções entre parâmetros de análise de variações.....	2
Tabela 2- Métodos tradicionais de pesquisa em transportes	13
Tabela 3 - Estudo da variabilidade com dados obtidos por métodos alternativos.....	28
Tabela 4 - Estudo da variabilidade com dados obtidos por métodos tradicionais.....	29
Tabela 5 - Dados fornecidos pelo MLS.....	32
Tabela 6 - Exemplo de atribuição Pt. 1	49
Tabela 7 - Exemplo de atribuição Pt. 2	49
Tabela 8 - Exemplo de agregação Pt. 1	51
Tabela 9 - Exemplo de agregação Pt. 2	51
Tabela 10 - Cálculo do Índice de Mobilidade	52
Tabela 11 – Descrição das APs do Rio de Janeiro	59
Tabela 12 - Atribuição de valores Pt. 1	62
Tabela 13 - Atribuição de APs para cada colaborador	62
Tabela 14 - Dias representativos	70
Tabela 15 - Dia útil representativo	72
Tabela 16 - Variação do IM entre sábado e dia útil.....	81
Tabela 17 - Teste t sábado – dia útil.....	73
Tabela 18 - Variação de IM entre domingo e dia útil.....	74
Tabela 19 - Teste t domingo - dia útil	75
Tabela 20 - Variação de IM entre 01/mai e dia útil.....	76
Tabela 21 - Teste t 01/mai - dia útil.....	76
Tabela 22 - Variação de IM entre 15/jun e dia útil.....	77
Tabela 23 - Teste t Corpus Christi - dia útil	78
Tabela 24 - Variação de IM entre 16/jun e dia útil.....	78
Tabela 25 - Teste t 16/jun - dia útil	79
Tabela 26 - Variação de IM entre feriados	80
Tabela 27 - Teste t entre feriados	80

LISTA DE GRÁFICOS

Gráfico 1 - Índice de mobilidade relativo para AP1.....	65
Gráfico 2 - Índice de mobilidade relativo para AP2.....	66
Gráfico 3 - Índice de mobilidade relativo para AP3.....	67
Gráfico 4 - Índice de mobilidade relativo para AP4.....	68
Gráfico 5 - Índice de mobilidade relativo para AP5.....	68
Gráfico 6 - Diferença entre índices de mobilidade - sab	73
Gráfico 7 - Diferença entre índices de mobilidade - dom	74
Gráfico 8 - Diferença entre índices de mobilidade – 01/mai.....	76
Gráfico 9 - Diferença entre índices de mobilidade – 15/jun.....	77
Gráfico 10 - Diferença entre índices de mobilidade – 16/jun.....	78

LISTA DE SIGLAS E ABREVIATURAS

Sab – Sábado

Dom – Domingo

Mai – Maio

Jun – Junho

MLS – Mozilla Location Service

IM – Índice de mobilidade

GPS – *Global Positioning System*

ERB – Estação Radio Base

AP – Área de Planejamento

LA – Location Area

CDR – Call Detail Record

API – Application Programming Interface

RA – Região Administrativa

PDTU – Plano Diretor de Transportes Urbanos

DNIT – Departamento Nacional de Infraestrutura de Transportes

1. INTRODUÇÃO

1.1. Contextualização

A análise e modelagem de dados em transportes é feita geralmente considerando um dia de análise. Isto é feito mesmo aceitando que o padrão de viagens dos indivíduos tenha alguma variação entre dias distintos (PAS E KOPPELMAN, 1986). Tal variabilidade ocorre devido ao fato de que as pessoas não tem a mesma necessidade todos os dias, por exemplo, compras em mercearia não é uma necessidade diária. Outros fatores que podem trazer variações na quantidade de viagens são eventos inesperados, tais como mudanças no tempo, alagamentos, falta de segurança, entre outros (SCHLICH E AXCHAUSEN, 2003).

Além disso, as necessidades e desejos individuais estão em constante mutação, assim como as condições do ambiente urbano, que diz respeito a infraestrutura e a performance dos modos de transporte das cidades. Isso faz com que os indivíduos modifiquem seus padrões de viagens de um dia para o outro. Algumas atividades, como por exemplo comer, dormir e trabalhar são repetidas todos os dias, mas outras como compras ou lazer não necessariamente são repetidos em uma base diária (DHARMOWIJOYO et al., 2015).

Supõe-se que o padrão de viagens consiste de rotina, pois entende-se que um indivíduo prefere repetir aqueles padrões que satisfizeram suas necessidades em ocasiões passadas. Esta hipótese é baseada na teoria de maximização da utilidade, que assume que pessoas tentam realizar atividades tão eficientemente quanto possível (SCHLICH E AXHAUSEN, 2003). Hanson e Huff (1981) citam que os indivíduos procuram maior satisfação ao invés de maior otimização, levando em conta que a otimização levaria à tomada de decisão constante, os indivíduos buscam manter uma rotina que o satisfizesse em um período anterior. Por isso, há indícios de que a maioria das pessoas estabelecem padrões de comportamento habituais no que tange à viagens.

Uma das principais distinções que pode ser feita quanto a variabilidade em viagens urbanas é quanto à dias úteis e finais de semana, denominada variação semanal pelo Manual de Estudos de Tráfego do DNIT. Ahas et al. (2010) disserta acerca do possível impacto das atividades dos finais de semana nas sextas e segundas, bem como a influência dos dias úteis ou o hábito de ir às compras na mobilidade dos sábados.

Ferraz e Torres (2004) definem índice de mobilidade como a quantidade de viagens urbanas realizadas. O índice de mobilidade é um dos elementos do padrão de viagens, entre os componentes estão a OD da viagem, a hora de partida, o tempo de

viagem e hora da chegada. Muitos destes componentes (Tabela 1) foram estudados nos mais diversos estudos coletados na bibliografia acerca do tema dessa dissertação. Este trabalho busca estudar o número de partidas diárias no intuito de calcular a variação do índice de mobilidade dentro da semana por meio de um estudo de caso na cidade do Rio de Janeiro.

Tabela 1 – Componentes do padrão de viagens

Origem e destino
Hora de partida
Tempo de viagem
Hora de chegada
Número de partidas diárias

O entendimento da variação da mobilidade é importante para projetar serviços de transporte, bem como entender como tal variabilidade afeta as características do pico diário (PAS E KOPPELMAN, 1986). Além disso, a validade e a confiabilidade do uso de apenas um dia em pesquisas de transporte depende do nível de variabilidade presente em um indivíduo entre dias distintos (HUFF E HANSON, 1986). Entende-se então que a variação tende a fornecer maior precisão às pesquisas em transportes. Além disso, o estudo deste tema pode aumentar a eficácia das políticas públicas de transporte, no sentido de propiciar melhor remanejamento do transporte público, como por exemplo, linhas temporárias de ônibus urbanos. Caceres et al (2007) citam que a precisão no mapa de mobilidade é ferramenta essencial para habilitar as autoridades administrativas a não apenas otimizar suas redes de transporte, mas também otimizar o investimento requerido na adaptação da infraestrutura para necessidades futuras.

Para o estudo da variação semanal da mobilidade é necessário a coleta de dados de dias distintos, os chamados dados longitudinais. Tais dados tem coleta que demandam muitos recursos financeiros e além disso podem ser vulneráveis a viés por seus respondentes não recordarem suas viagens com consistente precisão sobre um período de tempo prolongado (HANSON E HUFF, 1981). Entretanto, segundo Jones e Clarke (1988), o uso destes dados pode fornecer o mesmo nível de precisão no parâmetro estudado em um custo reduzido da coleta de dados, ou para um mesmo recurso, a variância no parâmetro estimado pode ser reduzido, comparado a uma amostra de um dia apenas.

Visando a redução do custo para a execução de pesquisas em transportes surgem métodos alternativos de coleta de dados, entre eles, existe o método baseado em telefonia móvel, que foi utilizado nesta dissertação. Tal método tem limitações e vantagens em comparação aos meios tradicionais de obtenção de dados, como por exemplo as pesquisas domiciliares e pesquisas na via.

Os métodos tradicionais de pesquisa em transporte acarretam em custo elevado aos administradores públicos. A estrutura necessária, bem como a quantidade de profissionais contribui para seu elevado custo. Além disso, apenas fornecem um instante do comportamento do tráfego, visto que não são realizados continuamente (QUEIROZ E RIBEIRO, 2017; WHITE E WELLS, 2002). Outra desvantagem bastante recorrente sobretudo em cidades em que o índice de criminalidade é alto, é que as pessoas podem ter razões para não reportar de maneira fiel as informações.

Calabrese et al. (2013) ainda lista outras três limitações dos métodos tradicionais, dando enfoque às pesquisas domiciliares. A primeira limitação é o uso de uma pequena amostra devido ao alto custo das pesquisas de transporte. Em seguida, a limitação espacial e temporal dos dados coletados, o que não possibilita ter um cronograma completo das viagens dos residentes na localidade, bem como variações sazonais. Por último, a baixa frequência de atualizações que limita a capacidade de resposta das políticas urbanas no que se relaciona ao rápido crescimento metropolitano e das mudanças socioeconômicas, de infraestrutura e variação do padrão de viagens.

Dentre as vantagens do método tradicional destaca-se o grande detalhamento obtido por tais pesquisas. A desagregação a nível de indivíduo fornece informações valiosas para a análise e modelagem de sistemas de transportes. A disponibilidade de atributos demográficos e socioeconômicos também tornam o prognóstico mais preciso e menos suscetível a erros.

O método de coleta de dados baseado em telefonia móvel tem potencial para uma grande amostra que pode ser obtida numa frequência diária ou até mesmo horária. Além disso, o fornecimento de dados não é prejudicado por condições meteorológicas de baixa visibilidade, como é o caso do Optical Character Recognition (OCR), e também não sofre deterioração do equipamento do sensor, tais como os detectores de loop e câmeras instalados ao longo das rodovias (CACERES et al., 2007)

As desvantagens deste método consiste na não aleatoriedade da amostra da população e no fato de que o conjunto de dados não foi projetado para propósito

referentes a transporte, conseqüentemente, o formato dos dados não vem de maneira fácil a ser utilizada, o que restringe a utilidade (CALABRESE et al., 2013).

Dentre os dados que podem ser extraídos por meio da telefonia móvel estão os dados de CDR (Call Detail Record) e os dados de sinalização. Tais dados são comumente utilizados em diversos estudos que utilizam mineração de dados no intuito de extrair informações acerca de transportes. Tais dados são de difícil obtenção, frequentemente faz-se necessário pagar para ter acesso aos mesmos. Existem poucos dados disponibilizados publicamente e gratuitamente com potencial para serem utilizados na área de transporte, dentre eles está o Mozilla Location Service.

Esta dissertação usou dados do Mozilla Location Service, uma plataforma que fornece dados locacionais publicamente e diariamente. Tais dados são provenientes dos colaboradores da plataforma, que alimentam a base de dados do MLS com informações referentes à antenas telefônicas e pontos de Wi-Fi públicos. Por questões de privacidade os dados acerca de localizações de Wi-Fi's não são liberados. Os dados de localização de antenas telefônicas são disponibilizados diariamente no site da plataforma, para uso público. O fato de ser um dado público traz grande relevância ao trabalho, haja vista que dados locacionais, geralmente, são de difícil obtenção, uma vez que esbarra em problemas de privacidade.

Este trabalho busca estudar a variação do índice de mobilidade entre dia útil, final de semana e feriado por meio do uso de dados gerados pelos colaboradores do Mozilla Location Service. O índice de mobilidade ou IM, é um indicativo da intensidade com que a rede de transporte é utilizada pelos colaboradores do Mozilla Location Service (MLS). Devido às características dos dados não foi possível estudar comportamentos dos indivíduos, o que levou à busca por um valor (índice) que agregasse todas as viagens realizadas pelos colaboradores no dia, o índice de mobilidade é o ideal para este caso.

A literatura abordou, até o momento, basicamente a variação do padrão de viagem dos indivíduos, isto é, mensurou-se a quantidade de rotina no comportamento individual. Quanto ao índice de mobilidade pouco se atentou. O estudo da variação do padrão de viagens pode trazer conclusões enganosas acerca do comportamento resultante de uma população. Suponha um exemplo, para um mesmo grupo será realizado um estudo acerca de variação, tanto do padrão de viagens, quanto do índice de mobilidade. Suponha também, que este grupo de indivíduos efetua viagens apenas nas segundas, quartas e sextas, com o mesmo destino e a mesma origem. Ao analisar a

variação do padrão de viagens a conclusão resultante seria de que quanto ao padrão de viagens os indivíduos tem baixa variabilidade, visto que eles vão sempre ao mesmo local às segundas, quartas e sextas, ou seja, não se consideram os dias que os indivíduos não viajaram. Entretanto, quando se considera o índice de mobilidade, a variação é alta visto que a quantidade de viagens difere dos dias em que não foi efetuada nenhuma viagem, terça e quinta, para os dias de segunda, quarta e sexta. Conclui-se que se tais indivíduos forem representativos à população, então a rede de transporte será menos utilizada nas terças e quintas, o que exige tratamento diferenciado de tais dias.

Outro exemplo importante a ser explanado é o da compensação que ocorre nos padrões de viagens. Supõe-se que um indivíduo realiza viagens diariamente para um dado local, entretanto, em determinado dia não foi possível realizar tal viagem. Esta viagem não realizada poderá ser compensada por outro indivíduo que, no mesmo dia fez um caminho diferente de sua rotina, mas semelhante ao caminho da pessoa que não realizou a viagem. Para considerar todas essas possibilidades deve-se considerar um comportamento agregado dos indivíduos, a fim de que quando todas as compensações forem realizadas, reste o impacto resultante na rede de transporte.

Esta agregação de viagens ainda não foi utilizada na literatura e tem o potencial de reduzir as compensações de padrões quando se estuda comportamentos individuais. O efeito desta compensação é que o impacto final, em determinada rede de transporte, pode ser nulo, embora haja variabilidade individual. Portanto, o estudo do coletivo pretende gerar conhecimentos acerca da resultante de todos os comportamentos individuais, a fim de fornecer parâmetros mais práticos para a tomada de decisão do administrador públicos, tal como, por exemplo, o remanejamento de uma linha de ônibus em determinado feriado para atender uma região com maior índice de mobilidade.

1.2. Problema de pesquisa

Qual o grau de variabilidade na mobilidade entre dia útil, final de semana e feriados para os colaboradores do Mozilla Location Service?

1.3. Hipóteses

- A variação do índice de mobilidade entre os dias de semana é significativa, sendo que sexta-feira é o dia de maior índice de mobilidade?

- O índice de mobilidade é reduzido igualmente no sábado e domingo em comparação ao dia útil?
- Existe variação significativa no índice de mobilidade entre feriados que caem em dias distintos da semana?

1.4. Objetivo geral

Estudar o grau de variabilidade do índice de mobilidade dos colaboradores do Mozilla Location Service entre dia útil, final de semana e feriados.

1.5. Objetivos específicos

- Identificar informações úteis aos estudos de transporte contidas no banco de dados do Mozilla Location Service (MLS)
- Elaborar um procedimento de mineração de informações acerca de variabilidade nos dados do MLS

1.6. Justificativa

A variação no que tange à mobilidade dos indivíduos existe inegavelmente. Estudar se essa variação ocorre em um grau significativo, ou não, tem grande importância e tem estimulado vários estudos ao longo dos anos.

Jones e Clarke (1988) destacam que se a variabilidade for melhor entendida, então o gerenciamento de sistemas de tráfego urbano pode claramente ser mais eficiente, assim como projetos relacionados a transporte podem ser combinados em uma maneira muito mais exata com o perfil da demanda. Schlich e Axhausen (2003) enfatizam que o perfil da demanda deve nortear a reorganização do sistema de tráfego, quando há a tentativa de produzir menor impacto ambiental. A oferta é organizada correspondendo às necessidades e desejos dos indivíduos, quanto mais complexo é seu comportamento, mais flexível a oferta deve ser (SCHLICH E AXHAUSEN, 2003)

O conhecimento aprimorado da variação na área de transportes pode contribuir para procedimentos de amostragem mais eficientes para contagens de tráfego em períodos curtos e também pode esclarecer qual o tempo mais apropriado para coleta de dados com vistas a modelagem. Além disso, pode haver o surgimento de uma nova família de modelos de alocação de tráfego para o método sequencial, no qual não

apenas o equilíbrio é buscado, mas também procura-se replicar a variabilidade existente entre indivíduos (JONES E CLARKE, 1988; HANSON E HUFF, 1988).

Kang e Scott (2010) argumentam que do ponto de vista de política de transporte, a análise da variabilidade em transportes pode refletir melhor as mudanças no padrão de comportamento dos indivíduos em resposta a políticas públicas de transporte quando comparado a estudos em dias únicos. Podem também propiciar remanejamento eficiente das rotas de transporte público.

Yang et al. (2016) destaca o aumento, nos últimos anos, da demanda de viagens nos feriados, os autores enfatizam que o estudo da variação em transportes se faz necessário a fim de mitigar os efeitos indesejados deste aumento, tais como congestionamentos e aumento no índice de acidentes. Susilo e Axhausen (2014) argumentam que a variabilidade fornece conhecimentos que ajudariam a melhorar o quadro de tomada de decisão aos planejadores das cidades.

Muitas pesquisas, em países como França, Reino Unido, Alemanha e Holanda já implementaram a coleta de dados de transporte em vários dias, essencial para o estudo da variabilidade (RAUX et al., 2016). Isso evidencia a importância crescente deste estudo ao entendimento da dinâmica do comportamento das viagens nas cidades.

1.7. Estrutura do trabalho

Este estudo prossegue com uma revisão da literatura no capítulo 2, neste capítulo serão abordados os métodos tradicionais e alternativos de coleta de dados em transporte. Além disso, no subitem 2.4 são elencados diversos trabalhos na literatura que discorrem acerca do tema, adotou-se uma divisão considerando o método de coleta de dados utilizado.

O capítulo 3 busca caracterizar os dados do Mozilla Location Service, para isso dissertou-se sobre os testes realizados e conclusões obtidas por meio destes testes. No capítulo 4 foi abordado o método proposto, tal capítulo contém o procedimento de tratamento e análise dos dados estudados. No capítulo 5 aplicou-se o método disposto no capítulo anterior na cidade do Rio de Janeiro.

No sexto capítulo estão dispostos os resultados obtidos por este trabalho. Por fim, no capítulo 7 foi realizada a conclusão do trabalho, assim como elencadas as recomendações de estudos futuros.

2. REVISÃO DE LITERATURA

Este capítulo trata-se de uma revisão, que contém conceitos e resultados de trabalhos referentes ao tema. Além disso, é feita uma explanação acerca do banco de dados utilizado neste estudo. O subcapítulo 2.1 é referente aos métodos tradicionais de coleta de dados em transportes, abordar tal assunto tem importância haja vista a possibilidade de comparação com os métodos alternativos utilizados não apenas por este trabalho, mas também por vários textos coletados na literatura. Em 2.2 é feita uma abordagem acerca dos métodos alternativos de obtenção de dados para transporte, é realizada uma breve revisão acerca dos dados provenientes de bilhetagem eletrônica e redes sociais e, por último, dada ênfase a telefonia móvel, estrutura que propicia a geração de dados utilizados neste estudo.

Os estudos acerca de variação em transportes vêm no subcapítulo seguinte. Parte imprescindível desta dissertação, estes estudos permitem a comparação de conclusões entre eles mesmos, bem como serão retomados nas discussões realizadas em relação às conclusões obtidas por este estudo. Visando maior organização e coesão do texto, optou-se por dividir tal subcapítulo em dois subitens, referentes à variação em transporte com dados obtidos por métodos tradicionais, no qual os dados são obtidos sobretudo por entrevistas, e variabilidade com dados de fontes alternativas. Neste último os métodos são diversos, englobando desde o uso de GPS, já recorrente nos estudos de transportes, até o uso de dados de redes sociais.

2.1. Métodos tradicionais de coleta de dados em transportes

Antes de apresentar os métodos tradicionais de coleta de dados em transporte, deve-se entender os dilemas envolvidos na escolha de um método ou outro. A escolha de um método é regida por três pilares, os quais são: recursos disponíveis, qualidade dos dados e quantidade dos dados (Figura 1). O dilema ou trade-off ocorre devido ao fato da impossibilidade de se controlar todos os três elementos principais. Desta maneira, considerando um dado recurso disponível, pode-se ou coletar uma grande quantidade de dados de baixa qualidade ou coletar uma limitada quantidade de dados de alta qualidade, sendo o último o procedimento mais adotado.

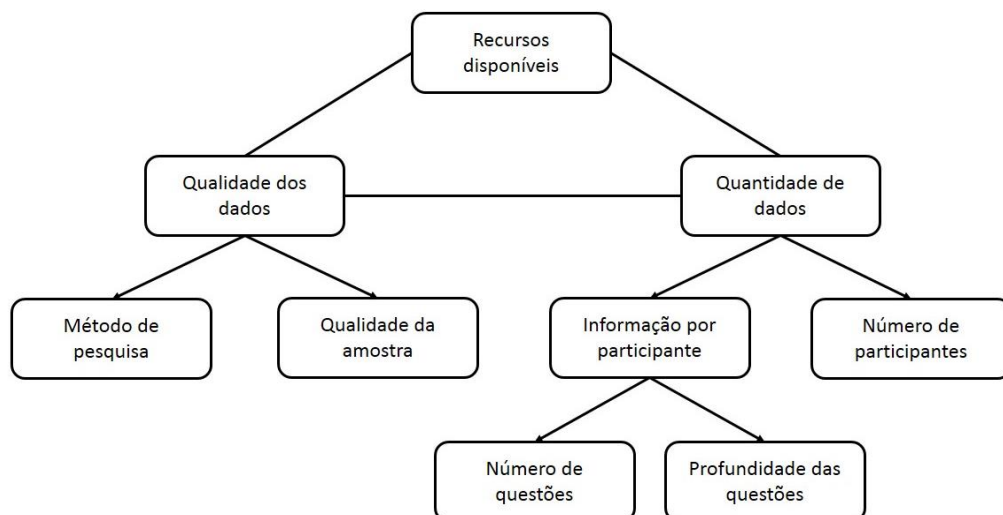


Figura 1 - Hierarquia de decisão para escolha do método de coleta de dados

Fonte: Adaptado de RICHARDSON et al. (1995)

A qualidade dos dados é função da qualidade da amostra obtida, assim como do método de pesquisa adotado. O método de pesquisa deve coletar dados da forma mais imparcial quanto possível, e isto deve ser monitorado por um controle de qualidade adequado ao padrão requerido aos dados coletados. Já a qualidade da amostra vai depender da capacidade da amostragem em ser representativa à população estudada, bem como dependerá também do quanto tal amostra foi obtida aleatoriamente.

A quantidade dos dados depende do dilema entre tamanho da amostra e quantidade de dados a serem recolhidos por cada integrante do conjunto amostral, isto é, por exemplo o número de perguntas em um questionário ou entrevista. Para um orçamento fixo e limitado ou são realizadas um pequeno questionário para um grande número de pessoas ou um questionário de alta profundidade a poucas pessoas.

Ainda no que toca a quantidade de dados coletado por participante há o trade-off entre o número de questões e a profundidade de tais questões, existem pesquisas que são efetivas na aplicação de um grande número de questões generalistas, já outras devem explorar tópicos com maior profundidade. Abordar tal dilema se deve ao fato de que a taxa de resposta geralmente diminui com o aumento do número de questões, tal taxa dependerá de fatores como o nível de interesse do tópico da pesquisa para os entrevistados e a qualidade geral do projeto do instrumento de pesquisa (RICHARDSON et al., 1995).

Os dilemas supracitados regem a escolha do método a ser utilizado para obtenção de informações não apenas na área de transportes, mas também nas ciências

humanas e sociais. Neste trabalho, serão abordados seis métodos de coleta de dados em transportes, tais tipos de pesquisa são: 1- Documentais, 2- Observacionais, 3- Auto resposta, 4- Por telefone, 5- Interceptação, 6- Domiciliares por entrevista.

Abaixo é feita uma explanação de cada um destes tipos de pesquisa, abordando suas definições, assim como suas vantagens e desvantagens.

2.1.1. Pesquisa documental

É realizada sob material inerte, inanimado, que frequentemente se resume a documentos, os quais não se espera nenhuma resposta destes objetos. Se resume na triagem em livros, artigos, mapas, fotografias ou bases de dados publicados ou não, na tentativa de descobrir a informação necessária ao estudo. Esta pesquisa é sempre realizada no início de qualquer estudo, então tem a vantagem de fornecer conhecimentos básicos acerca de um determinado tema, por exemplo, o Plano de Mobilidade de São Paulo fez uso de documentos municipais relativos à mobilidade, bem como legislações tanto estaduais, quanto federais.

Horários e mapas de rotas para a construção de um modelo de rede de transporte público e fotografias aéreas no estabelecimento de amostragem de casas em uma área de estudo são exemplos fornecidos por Richardson et al. (1995) do uso da pesquisa documental voltada ao transporte.

2.1.2. Pesquisa observacional

Pode ser realizada sob sujeitos inanimados ou animados, tal pesquisa se caracteriza também por não esperar nenhuma resposta específica do objeto estudado, apenas que ele se comporte de maneira comum quando estiverem sob observação. Na área de transporte, tem uso corriqueiro em estudos de tráfego. Podem validar outros estudos, mas não fornecem grande confiabilidade.

Contagem de tráfego de diferentes tipos, tais como em links, interseções, cordão ou screen line, ou pesquisas de tempo de viagem e atrasos em interseções são exemplos de pesquisas observacional na qual é obtida uma conclusão diretamente a partir da informação gerada pela coleta. Já quando se observam marcas de derrapagem para indicar locais perigosos numa rede rodoviária, não é possível apenas por estes dados gerar uma conclusão satisfatória, mas podem fornecer fortes indícios que validam conclusões obtidas por outros tipos de pesquisa. O Plano Diretor de Transportes Urbanos do Rio de Janeiro (PDTU-Rio) fez uso deste tipo de pesquisa nas contagens

volumétricas classificadas de tráfego, o estudo enfatiza que os dados obtidos tiveram grande importância na elaboração das matrizes de viagens, calibração da modelagem e identificação da variação dos fluxos ou sazonalidade.

2.1.3. Pesquisa de auto resposta

Muito utilizada atualmente sobretudo via e-mail, as pesquisas de auto resposta deixam a cargo do respondente a tarefa de ler e compreender, formular a resposta mentalmente e transcrever em um formulário eletrônico ou físico. Diversos estudos utilizam tal método por ser de grande alcance, visto que boa parte da população tem acesso ao e-mail de maneira relativamente rápida, tal como em um smartphone. Deve-se atentar ao fato de que, dependendo do objeto de estudo, tais dados podem ter grande viés, haja vista que apenas pessoas com algum nível de conhecimento digital irão completar satisfatoriamente o questionário.

Além disso, é necessário observar a baixa taxa de resposta de tais métodos. Mesmo com sendo um método de grande abrangência por baixo custo, uma baixa taxa de resposta pode inviabilizar a coleta.

2.1.4. Pesquisa por telefone

Esta pesquisa é caracterizada pelo não contato visual entre o entrevistador e o entrevistado. As vantagens são basicamente as mesmas das pesquisas de auto resposta, isto é, a alta cobertura geográfica e a possibilidade de coletar dados de maneira mais barata. Entretanto, as perguntas devem ser realizadas em um tempo relativamente curto, uma vez que, segundo Richardson et al. (1995), a taxa de resposta cai a partir de 10 minutos de questionário.

2.1.5. Pesquisa de interceptação

São realizadas fora do domicílio do respondente, enquanto o mesmo está em algum modo de transporte ou realizando uma atividade rotineira, como por exemplo, fazer compras. Tal método é essencial em estudos acerca do grau de satisfação do usuário de algum serviço de transporte, entretanto, para cada modo deve-se projetar o número e profundidade de questões adequado. Por exemplo, não é apropriado um questionário extenso em linhas de ônibus caracterizadas por um intenso sobe e desce, já os passageiros de voos não se incomodariam em responder um grande número de

perguntas, haja vista o caráter de recreação que a pesquisa ganha devido a característica do modo aéreo.

O método também é essencial em pesquisas origem-destino. Além do pesquisador, a interceptação comumente é feita com o apoio de uma autoridade policial, a fim de propiciar segurança ao respondente. As informações coletadas pelo PDTU-Rio em pesquisas de interceptação foram relativas a origem-destino dos veículos, ocupação, motivo da viagem e local da residência.

2.1.6. Pesquisa domiciliar por entrevista

Nos estudos acerca de variabilidade é o método tradicional de maior uso, é direto e há grande interação entre o entrevistado e o entrevistador, fato que deve ser evitado, na medida do possível, a fim de evitar distorções. Richardson et al. (1995) cita que embora essa interação possa permitir coletar dados mais complexos, também permite um maior grau de viés nas respostas. Possui maior taxa de resposta quando comparada a outros métodos de pesquisa, bem como há a possibilidade de obtenção de dados mais detalhados e completos. Entretanto, a desvantagem é que são mais caras de se empreender.

O PDTU-Rio fez uso deste método de pesquisa, o plano admite a dificuldade inerente ao método, haja vista que deve-se convencer os indivíduos a interromperem suas rotinas e responderem questões relativas a suas vidas. Entretanto, diversos procedimentos são adotados para que a pesquisa se torne eficaz, resultando no tipo de pesquisa que retorna os dados de grande profundidade, tais como indicadores sociais e deslocamentos realizados no dia anterior à entrevista.

A Tabela 2 resume os métodos tradicionais de pesquisa em transportes abordados por este estudo.

Tabela 2- Métodos tradicionais de pesquisa em transportes

Tipo de pesquisa	Resumo	Vantagens	Desvantagens
Documentais	Material inanimado, documentos, não se espera resposta	Fornecimento de dados básicos	Depende fortemente do procedimento de coleta
Observacionais	Inanimado ou não, não se espera resposta	Validam outros tipos de pesquisa	Não permitem alto grau de confiabilidade
<i>Auto resposta</i>	Encaminhado pelo pesquisador, respondido e enviado pelo respondente	Baixo custo Rapidez Flexibilidade do horário para responder	Baixa taxa de resposta Dúvidas podem gerar dados incompletos
Por telefone	Executadas ao telefone	Alta cobertura geográfica Possibilidade de supervisão do entrevistador	Limite de tempo para a pesquisa (taxa de resposta cai a partir de 10min) Número de pessoas entrevistadas em um domicílio é limitada a 1
Interceptação	Pesquisas fora do domicílio, em um modo de transporte ou atividade rotineira	Dependendo do tipo podem ser de baixo custo	Incapacidade de acompanhar ou inferir comportamentos de não respondentes
Entrevista	Interação efetiva entre o entrevistador e entrevistado	Maiores taxas de resposta Flexibilidade O entrevistador pode ser efetivo para manter o interesse do entrevistado	Mais caras Clusterização dos domicílios pode reduzir a acurácia dos dados Viés provocado pelo entrevistados

2.2. Métodos alternativos de coleta de dados em transporte

Os métodos alternativos de coleta de dados nasceram do desenvolvimento tecnológico de diversas áreas, dentre as quais se destaca a telefonia móvel. Tais métodos vem ganhando importância, uma vez que os métodos tradicionais, no geral, tem execução cara e demorada. Além de serem mais baratos, os métodos alternativos

usualmente são coletados de maneira passiva, o que minimiza o problema relacionado a confiança na memória dos participantes (Bekhor et al., 2013).

Feriancic et al. (2015) destaca que as pesquisas tradicionais também falham na coleta de informações de diversos pontos importantes de uma cidade, locais que, ainda que não gerem uma quantidade de deslocamentos comparáveis aos propósitos de trabalho e estudo, são de grande concentração de viagens do tipo saúde, compras e lazer, e crescem em importância por concentrar parte da demanda de viagens. Métodos alternativos possibilitam o estudo de pontos geradores de tráfego, sejam eles temporários ou permanentes, como grandes eventos (Asakura et al., 2005; Sagl et al., 2012) ou estações de metrô.

A possibilidade de coletar dados em tempo real, outra característica de grande parte dos métodos abordados neste subcapítulo, reduz o tempo de execução das pesquisas, fazendo com que se tenha uma matriz OD diária e dinâmica, ao invés da frequência comum, 10 anos. Os métodos alternativos tem grande dinamicidade, a cada surgimento de dados locais, independentemente da fonte, há esforços para utilizá-los na área de transportes. Devido a relativa pouca idade dos estudos, poucos conhecimentos estão consolidados, fazendo com que essa fronteira seja muito promissora.

Nos próximos itens é feita uma revisão dos métodos alternativos, dando enfoque à telefonia móvel, dado utilizado nesta pesquisa.

2.2.1. Estudos voltados ao transporte público

Os métodos voltados ao transporte público baseiam-se principalmente em bilhetagem eletrônica. O Sistema de Bilhetagem Eletrônica (SBE) surgiu com os principais objetivos de coibir fraudes advindas de liberações indevidas de gratuidade, redução do tempo de filas para embarque e otimizar os recursos destinados à venda de passagens. O cartão, principal elemento da bilhetagem, permite a coleta contínua de dados transacionais individualizados sobre o uso de sistemas de transporte (BRITO, 2016).

Os dados gerados pelo SBE foram utilizados por Guerra et al. (2015) em um estudo que busca estimar a matriz Embarque-Desembarque para o sistema de transporte público para a cidade de Maceió, Alagoas. Brito (2016) utilizou o mesmo tipo de dado para elaborar um método que determinou a ocupação dos trens nas viagens do sistema

ferroviário na cidade do Rio de Janeiro, o método possibilitou a identificação de viagens com ocupação superior ao recomendado, bem como composições superdimensionadas.

Os estudos relacionados à bilhetagem eletrônica são a maioria no que toca à coleta de dados alternativos relativo ao transporte público. Além desses, cabe salientar uma metodologia elaborada por Dos Santos Melo et al. (2015), os autores elaboraram um sistema de comunicação baseado na tecnologia Bluetooth. Bluetooth é um padrão que especifica a comunicação entre dispositivos a curta distância via rádio, com baixo consumo de energia e baixo custo no estabelecimento de conexões (DOS SANTOS MELO, 2015). Esta tecnologia também dispensa o uso de dados móveis de telefonia, o que acarreta em economia ao usuário do sistema.

Além dos estudos citados acima, enfatiza-se que os métodos de coleta de dados provenientes de telefonia móvel e redes sociais também são utilizados na área de transporte público. A aplicabilidade dos dados de telefonia móvel depende da eficiência em lidar com os vários modos de transporte possíveis em que o dispositivo pode estar, existem métodos capazes de estimar o modo de transporte com base na velocidade média, e a partir disso analisar viagens de ônibus, por exemplo, em que se supõe que a velocidade média é menor que nos carros. O mesmo problema ocorre com dados provenientes de redes sociais, entretanto, estes estudos geralmente analisam dados de check-in que fornecem a localização de estações, pontos de ônibus entre outros pontos que o usuário optar por compartilhar.

2.2.2. Estudos baseados em redes sociais

Redes sociais são muito populares atualmente, existem em grande número e cada qual com sua característica, desde a postagem de textos curtos, tal como o Twitter, passando por uma rede de conteúdo geral, como o Facebook ou Google +, até redes especificamente de fotos e vídeos, são exemplos o Instagram e o Flickr. Estes serviços permitem que usuários se conectem com amigos e tenham a capacidade de gerar informações temporais e espaciais associadas a postagem.

Os dados provenientes de redes sociais também podem ser obtidos sem interrupções, visto que são gerados como um subproduto da interação com a plataforma. Além disso, estão disponíveis em Big Data, conceito que significa o conjunto massivo de dados sendo produzido de maneira contínua, e esta continuidade permite o estudo de mudanças em tempo real (McNeill et al., 2016).

Sparks et al. (2016) cita que a vantagem de utilização de tais dados é a natureza que os mesmos têm de serem em tempo real e também a capacidade de fornecer dados não apenas de onde algo ou alguém está mas, possivelmente, o propósito de estarem em determinado local por meio das descrições textuais. Baseando-se nisso, Sparks et al. (2016) utilizaram técnicas de mineração de textos e aprendizagem da máquina para interpretar *tweets* georreferenciados para determinar o local onde o indivíduo está com base no texto enviado. O principal objetivo dos autores foi tentar incorporar dados não explicitamente vinculados aos *tweets*, o que pode aumentar significativamente a quantidade de dados disponíveis.

Mcneill et al. (2016) também utilizaram dados do Twitter para estudar uso do solo. A suposição feita pelos autores foi de que os *tweets* enviados durante as horas de trabalho do dia são mais prováveis de serem de um local de trabalho, o mesmo vale para o horário de casa. Os trabalhos dispostos neste item ilustram brevemente como alguns dados proveniente de redes sociais estão sendo utilizados no estudo em transportes, a principal dificuldade, assim como em basicamente todos os métodos alternativos, é o acesso aos dados, uma vez que devido a questões de privacidade, tais dados tem liberação limitada. O Twitter é uma rede notável quanto a isso, uma vez que ela fornece os dados publicamente para pesquisa, por este motivo, na literatura grande parte dos estudos utilizam dados desta rede social.

2.2.3. Estudos baseados em telefonia móvel

Embora seja distinto dos dados utilizados pelos estudos elencados neste item, os dados do MLS são gerados com base na mesma rede e obedecendo alguns padrões citados aqui, por isso, será dado maior enfoque neste item. Os aparelhos celulares e as ERBs (Estações Radio-Base) são os elementos do sistema com maior proximidade ao usuário, sistema que ainda conta com a Central de Comutação e Controle (CCC), elemento responsável por gerenciar as diversas peculiaridades do sistema, tal como o *handover* que é a transferência de EBRs quando o aparelho está em ligação. O CCC gerencia a mudança de célula de maneira imperceptível ao usuário.

Cada Estação Rádio Base (ERB) atende a uma certa região, essa região denomina-se Célula. As células podem ser omnidirecionais ou setORIZADAS, conforme figuras 2 e 3 abaixo.

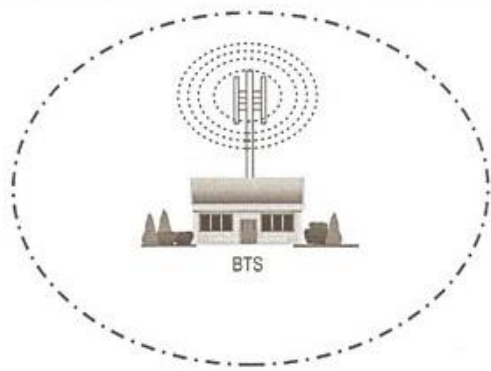


Figura 2 - Células omnidirecionais

Fonte: Santos (2008)

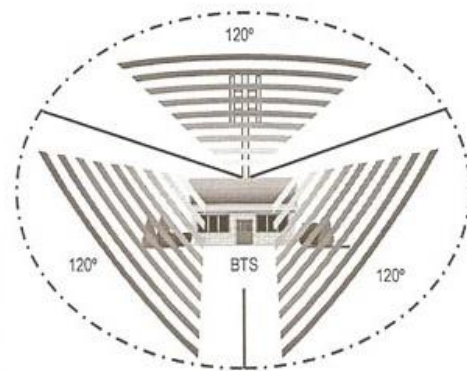


Figura 3 - Células setorizadas

Fonte: Santos (2008)

No primeiro caso, existe uma única ERB que atende uma região em 360° , sendo mais utilizadas em regiões de baixa demanda, como em zonas rurais por exemplo. Já as células setorizadas irradiam sinal conforme a direção estabelecida, geralmente são divididas em ângulos de 120° , tal acerto permite alta demanda, sendo mais utilizado em grandes centros. O conjunto composto por torre e ERB é popularmente conhecido como antena.

As antenas são agrupadas para formar as *Location Area (LA)*, que existem por razões relacionadas com a gestão da rede de telefonia, uma vez que isto torna possível identificar mais rapidamente telefones móveis no caso de uma chamada ou SMS. A LA de cada telefone é conhecida a todo momento, enquanto que a posição em nível de antena é conhecida apenas quando é realizado uma ligação (BONNEL et al., 2015).

A área de cobertura de uma antena varia conforme a região de serviço, o que impacta diretamente na precisão dos métodos de localização baseados em telefonia. Conforme Caceres et al. (2007) a área de cobertura, e por conseguinte a precisão, pode ser de 200m em áreas urbanas até 5 a 20km em áreas rurais. Dessa maneira, para um estudo com nível de confiabilidade aceitável deve-se trabalhar em áreas urbanas densamente ocupadas, ou então de acordo com Wang et al. (2013), os dados tem uso adequado para viagens de longa distância ou viagens interurbanas.

Além da imprecisão intrínseca à estrutura do sistema de telefonia móvel, há ainda o problema de sobrecarga em células, o que pode aumentar os erros de localização. Em células sobrecarregadas, em períodos de pico de uso, os usuário são automaticamente transferidos de uma rede mais próxima para um rede mais distante, mas menos carregada (WANG et al. 2013). Calabrese et al. (2011) cita que esta imprecisão limita o tamanho das regiões a serem consideradas em um estudo.

A estrutura explicada nos parágrafos anteriores geram dois tipos de dados locais úteis ao estudo em transportes, o *Call Detail Record* (CDR) e o dado de rastreamento. O CDR é automaticamente recolhido pela operadora para necessidades de faturação, são gerados pelas atividades de comunicação, tais como chamada, SMS ou conexão à internet de cada telefone celular. São registradas no formato id – tipo de atividade – tempo – localização. Os dados de rastreamento registram os dados que passam através da antena. Além do próprio CDR, *handovers*, atualização de LA e registros de quando o telefone é ligado ou desligado. Geralmente, a resolução da localização de CDR e rastreamento está limitada à localização da antena. No entanto, a precisão da localização pode ser melhorada utilizando o método de triangulação ou intensidade do sinal (Raimond e Couroneé, 2014; Bonnel et al., 2015).

Bekhor et al. (2013) utilizou dados de telefonia a fim de estudar padrões de viagens de longa distância. O estudo foi feito em nível nacional, utilizando dados de rastreamento. Bonnel et al. (2015) também fez uso de dados de CDR e rastreamento para estimar a matriz OD para a cidade de Paris. A validação realizada pelos autores resultou em ótima correlação com o *Enquête Globale Transport*, a mais completa pesquisa relacionada a área de transportes da capital da França.

Os dados do MLS apresentam semelhanças e distinções com relação aos dados apresentados aqui. Fazem uso da mesma estrutura, trazem os mesmos problemas de precisão, entretanto, não são coletados de maneira passiva, tais como o CDR ou dados de rastreamento, ou seja, nos dados utilizados por este trabalho é necessária a participação ativa do usuário do aplicativo Mozilla Stumbler, denominado colaborador, que coleta os dados de localização para o MLS de maneira voluntária.

No capítulo posterior é feita uma revisão do tema principal desta dissertação.

2.3. Variabilidade em transportes

A modelagem da demanda de viagens geralmente é feita considerando 1 ou 2 dias de dados coletados por entrevistas em domicílio. Este tipo de dado é conhecido como dado de corte transversal, que consiste em uma amostra de uma unidade de análise, tomada em um determinado ponto no tempo (AMARAL, 2010). O conjunto de dados acerca de viagens é obtido por meio da solicitação que cada familiar do domicílio da amostra relate seu comportamento em um dia escolhido aleatoriamente na semana (PAS, 1987).

A análise de dados transversais assume que as viagens dos indivíduos são repetitivas e de natureza habitual (STOPHER E ZHANG, 2011). A base para tal consideração vem da teoria de maximização da utilidade, que afirma que indivíduos fazem escolhas para maximizar seu bem-estar. No geral, os indivíduos tentam meios convenientes de executar uma atividade ao invés de procurar a maneira ótima, pois um comportamento otimizador levaria a frequente tomada de decisão, já o comportamento rotineiro, que reduz a quantidade de análises e julgamentos, minimiza o estresse, o que tende a aumentar o bem-estar. Com isso espera-se que as pessoas estabeleçam um comportamento habitual no que tange a viagens (AMARAL, 2010; ELANGO et al., 2007; HUFF E HANSON, 1981).

Segundo Pas (1987), ainda que se reconheça a existência da variabilidade diária no padrão de viagens, o uso de dados de seção transversal é uma prática aceita. Entretanto, esta abordagem é criticada por falhar na explicação e previsão das reais necessidades e demandas individuais (DHARMOWIJOYO et al., 2015b). Zhong et al. (2015) destaca que o estudo da variabilidade é tão importante quanto o estudo da regularidade, uma vez que variabilidade identifica diversidade. Entender o grau de variabilidade é compreender de uma melhor forma a dinâmica do sistema de transporte das cidades.

A variabilidade é inerente à característica do indivíduo que reside nas cidades da atualidade. Obrigações relativas a trabalho ou estudo tendem a moldar um comportamento rotineiro e repetitivo, por outro lado, limitações espaciais, temporais, necessidades, atividades físicas ou sociais tendem a aumentar a variabilidade no padrão de viagem dos indivíduos.

Para identificar a variabilidade é necessário o uso de dados longitudinais, que ao contrário dos dados de corte transversal, possibilitam repetidas observações ao longo do tempo. Estes dados fornecem informações valiosas para análise e para estimação de mudanças microscópicas que ocorrem no nível de domicílio ou de indivíduo, além disso, podem facilitar a identificação da relação causa e efeito devido ao fato de que levam em conta a dinâmica de comportamento (ELANGO et al., 2007).

A coleta de dados longitudinais para transporte é mais cara comparada a dados de corte transversal. Dependendo do método de coleta podem ser vulneráveis a viés, uma vez que os respondentes podem não recordar seus comportamentos com precisão suficiente por um período de tempo prolongado. Também deve-se atentar ao fato do

fardo causado às pessoas entrevistadas, o que reduz a adesão a tais pesquisas (ELANGO et al., 2007; HUFF E HANSON, 1981; SCHLICH E AXHAUSEN, 2003).

As observações dos padrões de viagens dos indivíduos ao longo do tempo permite mensurar a variabilidade interpessoal e intrapessoal. A variabilidade interpessoal é a diferença entre o padrão de viagem de diferentes indivíduos, no mesmo dia ou em dias diferentes, enquanto que a variabilidade intrapessoal refere-se a diferenças nos padrões de um mesmo indivíduo sobre um determinado tempo (PAS E SUNDAR, 1995; KANG E SCOTT, 2010).

Na literatura, há uma vasta gama de dados utilizados para o estudo da variabilidade, variando entre dados com pequena amostra de grande profundidade de informação e dados de grande amostra com informações mais gerais, frequentemente apenas informações locais. Uma grande profundidade das informações é obtida, geralmente, por métodos tradicionais de pesquisa em transporte, já as massivas informações locais são coletadas por GPS (*Global Positioning System*) ou CDR (*Call Detail Record*) da telefonia móvel.

O subcapítulo 2.5 é referente a uma revisão de literatura sobre o objeto de estudo desta dissertação, dando enfoque aos dados utilizados e resultados obtidos no que toca à variabilidade entre dia útil, final de semana e feriados.

2.4. Estudos acerca da variabilidade da mobilidade

Nos trabalhos referentes à variabilidade, percebe-se a existência de dois grandes grupos no que se refere aos dados utilizados. No primeiro grupo estão os trabalhos que utilizaram métodos tradicionais, tal como a entrevista domiciliar, para efetuar a coleta. O segundo grupo é referente a estudos que utilizaram métodos alternativos de coleta de dados, tais como a estrutura da telefonia móvel ou GPS. Tais grupos nortearam a subdivisão dos subitens seguintes.

2.4.1. Variabilidade com dados obtidos por métodos tradicionais

Grande parte dos estudos voltados à variabilidade tiveram início nos anos 80, são trabalhos que construíram o conhecimento básico acerca do tema, visto que, anteriormente, pouco se tratava sobre o assunto. Hanson e Huff (1981) fizeram uso do *Uppsala Household Travel Survey*, conjunto de dados coletados por meio de formulários de auto resposta em um período de 35 dias consecutivos, capaz de capturar a variação diária no comportamento de viagem do indivíduo. Os autores mostraram que, embora

haja alto grau de repetição, há grande quantidade de variabilidade intrapessoal mesmo em dias de semana, ou seja, os padrões para um indivíduo diferem consideravelmente de um dia para o outro.

Huff e Hanson (1986) concluíram que as observações de um dia simples por indivíduo não são representativas da gama de padrões exibidas por uma pessoa em um período mais extenso, o que leva a rejeitar a visão de que viagens são altamente repetitivas. Os próprios autores destacam que o principal achado é que os indivíduos exibem mais de um padrão diário e que estes padrões são fundamentalmente diferentes um do outro. Huff e Hanson (1988) recomendam o uso de uma semana, no mínimo, para a coleta de dados a fim de observar os diferentes padrões de viagens de um indivíduo, no entanto, salientam que não é tempo suficiente para observar a frequência com que tais padrões ocorrem.

Pas e Koppelman (1986) utilizaram o *Reading Activity Diary Survey*, que contém uma espécie de diário do indivíduo, bem como um questionário feito mediante a presença de um entrevistador. Com uma amostra de 112 indivíduos em um período requerido de sete dias por participante, tal pesquisa foi realizada em 1973. Tais dados propiciaram um nível notável de detalhamento socioeconômico e de viagens, o que possibilitou análises além dos fatores locais. Os autores descobriram que pessoas com emprego tem um nível de variabilidade intrapessoal muito menor, no que toca a frequência de viagens, quando comparado a pessoas desempregadas, ou seja, os padrões de viagens de pessoas empregadas são mais rotineiros e repetitivos que de pessoas desempregadas. Além disso, os autores descobriram que a classe social e a disponibilidade de recursos para se efetuar uma viagem afetam a variabilidade intrapessoal do subgrupo referentes às pessoas desempregadas.

Pas (1988) descobriu que, para os dados da *Reading Activity Diary Survey*, há uma tendência da amostra em fazer padrões mais complexos de viagens nas quintas-feiras. Ademais, constatou que padrões semanais de viagens podem ser agrupados dentro de um pequeno número de padrões-tipo retendo a maior parte da informação de padrões originais, ou seja, a variabilidade não é tão significativa. Os resultados também indicaram que os padrões em dias úteis são independentes do dia da semana e que tais dias da semana diferem muito pouco entre si, no que toca a padrão de viagens.

Pas e Sundar (1995) utilizaram dados longitudinais obtidos por meio da Prefeitura de Seattle, no ano de 1989. Os resultados indicam que a variabilidade intrapessoal explica uma fração considerável da variabilidade nos aspectos do

comportamento diário de viagens. Entende-se, então, que para estudo da variabilidade o uso de dados por indivíduo acerca da variação de mobilidade é mais apropriado

O *Mobidrive* é um outro conjunto de dados longitudinais que sustentou vários estudos relacionados a variabilidade. A pesquisa durou um total de seis semanas, na qual cada respondente tinha de preencher um diário relativo às viagens feitas no dia. Um total de 317 indivíduos de 139 domicílios participaram da pesquisa. Utilizando o *Mobidrive* Schlich e Axhausen (2003) notaram que a similaridade declina se o método capturar a maior complexidade do padrão de viagens, ou seja, dependendo do método de análise, para o mesmo dado pode-se mensurar variabilidades distintas. Tal afirmação tem relativa consonância com Jones e Clarke (1988) que dissertam que a existe uma grande estabilidade se a análise for realizada no nível agregado do padrão de viagens, o que frequentemente compensa mudanças no comportamento a nível micro.

Schlich e Axhausen (2003) ainda destacam que o padrão de viagem é claramente mais estável em dias úteis do que em finais de semana, havendo baixa correlação entre o comportamento observado nestes dois períodos. Além disso, há destaque para a alta correlação entre sexta-feira e outros dias de semana, o que mostra que não há grandes diferenças para os diferentes dias da semana no que concerne ao padrão de viagens. Os autores ainda refutam a assertiva de Huff e Hanson (1988), pois concluiu-se que o período mínimo para coleta de dados, se o objetivo for a mensuração da variabilidade, é de duas semanas.

Ainda utilizando o *Mobidrive* Susilo e Kitamura (2005) descobriram que sextas-feiras fogem do padrão de mobilidade da semana e se assemelham ao sábado, o que contraria Schlich e Axhausen (2003), já domingo tem menor intensidade de mobilidade em comparação com todos os outros dias, isto é, as pessoas ficam mais em casa. Os autores também destacam que, embora haja variabilidade, nos dias úteis há um padrão visivelmente recorrente, ao contrário dos finais de semana que se caracterizam pela randomicidade. Observa-se até aqui que para os mesmos dados há a possibilidade de resultados distintos, dependendo da metodologia utilizada pelos autores.

Tarigan e Kitamura (2009) buscaram estudar viagens com o propósito de lazer utilizando um dado longitudinal de seis semanas. Indivíduos decidem executar atividades de lazer com base em experiência positiva obtida de viagens a lazer anteriores e então simplesmente repetem viagens semelhantes sem julgar a utilidade (bem-estar) novamente (TARIGAN E KITAMURA, 2009). Portanto, o princípio que rege as viagens relacionados ao propósito de lazer é similar ao princípio de

maximização da utilidade, que é intrínseco ao comportamento do indivíduo para os mais variados propósitos de viagens.

Entretanto, os autores acharam diversas diferenças entre as viagens a trabalho em comparação com viagens a lazer. A frequência média de viagens para trabalho por semana é muito maior que a frequência média de viagens por semana para atividades de lazer. Entretanto, a variabilidade no número de viagens por semana para trabalho é muito menor que o número de atividades de lazer, isto é, as viagens pendulares ao trabalho fazem com que o padrão calculado seja predominantemente repetitivo em dias úteis, já o lazer é mais imprevisível ou com menor padrão observável, conforme salienta abaixo.

As necessidades e desejos em viagens para fins de lazer são mais variáveis e complexas do que para outros fins. Para viagens a trabalho é provável que os passageiros tenham padrão de viagem estável ao longo do tempo, pois isto é visto como uma atividade obrigatória, em vez de ser conduzido pelos desejos e necessidades dos indivíduos (TARIGAN E KITAMURA, 2009).

Ahas et al. (2010) coletaram dados por meio de questionário submetidos a 576 respondentes na área metropolitana de Taillin, Estônia. Os autores acharam que sábado durante a tarde é o movimento típico para atividades de lazer e compras em localidades próximas ao centro da cidade. Também estudaram a variação da intensidade de mobilidade e descobriram que 10% da amostra permanecem em casa durante o fim de semana, apenas efetuando pequenos deslocamentos. Por fim, afirmam que dias úteis tem ritmo similar, e que indivíduos que viajam pendularmente parecem permanecer em casa nos fins de semana.

Kang e Scott (2010) fizeram uso de uma pesquisa de auto resposta denominada *Toronto Travel-Activity Panel Survey* (TTAPS). Os autores observaram que, em dias de semana, padrões de viagens não são uniformes de um dia para o próximo e um dia típico não foi constatado. Esta descoberta, segundo os autores, voltam a atenção a confiabilidade de estudos acerca de demanda de viagens, que usam principalmente um conjunto de dados semanais agrupados para criar um dia típico.

Dharmowijoyo et al. (2015b) alcançou a maior profundidade de detalhamento a nível de indivíduo dentro os estudos contidos na revisão desta dissertação. Além das características socioeconômicas e relacionadas quantidade de viagens, comumente coletadas por diversos estudos, também foi realizada a coleta de dados acerca de

atitudes dos indivíduos concernentes ao uso de diferentes modos de transportes, bem como a sociabilidade, frequência de atividades físicas e estilo de vida do indivíduo, isto é, questões relacionadas a alimentação, fumo, álcool, etc.

Dentro os tantos resultados alcançados pelos autores será dado enfoque aqui à diferença existente entre o número de viagens médio entre a sexta-feira e os demais dias da semana (segunda a quinta), 2,76 e 2,52, respectivamente. Entende-se então que os indivíduos da área de estudo, no caso Jacarta (Indonésia), efetuam mais viagens na sexta-feira do que em outros dias úteis, aspectos relacionados à socialização ou lazer podem explicar tal diferença. Já nos finais de semana, o número médio de viagens cai (2,29) o que indica que a intensidade da mobilidade se reduz, ou em outras palavras, os indivíduos habitualmente ficam em casa nos finais de semana.

Dharmowijoyo et al. (2017), com os mesmos dados citados anteriormente, cita que a variabilidade entre finais de semana é baixa, o mesmo é observado entre dias úteis. Dharmowijoyo et al. (2015) fez uso do conjunto de dados do *SIMTRAMP Household Travel Survey*, pesquisa realizada também na cidade de Jacarta. Os dados são referentes ao diário de viagens de 3846 respondentes e 1175 domicílios por duas semanas consecutivas. Os autores acharam que sexta-feira tem o mais alto número de viagens, no domingo, entretanto, indivíduos tendem a ter menor número de viagens. Provavelmente, conforme os autores, porque domingo, diferentemente de sábado, é um dia voltado completamente ao descanso, no qual o indivíduo tende a passar mais tempo em atividades domésticas. Os dados obtidos possibilitaram analisar, inclusive, que os cronogramas são mais restritos em dias úteis e mais flexíveis durante os dias de fins de semana.

No intuito de estudar deslocamentos em feriados Yang et al. (2016) analisaram duas fontes de dados contendo características socioeconômicas e informação sobre as viagens, tais como local de partida e destino, hora de início e término, propósito da viagem e modo de transporte. Os conjuntos de dados utilizados foram o *Beijing Comprehensive Travel Survey* (BCTS) e o *Beijing Holiday Travel Survey* (BHTS). Os autores não estabelecem conclusões acerca da intensidade da mobilidade em feriados, entretanto, concluem que a estrutura de decisão em dias úteis é diferente dos feriados. Em dias úteis os indivíduos priorizam o roteiro (origem e destino) de viagens em detrimento do modo de transporte, sendo o inverso em feriado. Então, pode-se deduzir que um sistema de transporte público deficiente não afetaria as viagens pendulares relativas a trabalho ou estudo, no entanto, reduziria a intensidade das viagens em

feriados. Yang et al. (2016) ainda concluem que em feriados os indivíduos têm maior dependência dos carros que em dias de semana.

Raux et al. (2017) coletou dados por um período de uma semana para a cidade de Gante, Bélgica, foram obtidos 717 diários de viagens e atividades. Os dados possibilitaram a análise por propósito e percebeu-se que o número médio de viagens com o propósito de trabalho em dias úteis é quatro vezes maior que nos sábados e 13 vezes maior que no domingo. Quanto ao propósito de estudo a redução chega a 14 vezes, tanto no sábado quanto no domingo. Tal resultado leva a crer que mesmo sábado sendo um dia no final de semana, muitos indivíduos trabalham ainda que por um período reduzido em comparação com dias úteis. Já domingo se configura como o dia de menor intensidade de mobilidade, exceto para os propósitos de visita a familiares ou viagens a lazer.

Nota-se, com esta primeira parte da literatura acerca do objeto de estudo, que existem divergências e semelhanças entre as conclusões obtidas pelos diversos trabalhos citados nesta revisão. Os métodos de análise, bem como os dados coletados influenciam sobremaneira nos resultados, além do fato de as cidades abrigarem características distintas o que dificulta generalizações. Na busca por métodos mais baratos e que tragam resultados mais confiáveis acerca dos padrões de viagens, há uma crescente busca por métodos alternativos de coleta de dados. No subitem seguinte é feita uma revisão referente aos trabalhos que usam tais métodos.

2.4.2. Variabilidade com dados obtidos por métodos alternativos

Os métodos alternativos se referem a tecnologias emergentes na área de transporte. As tecnologias abordadas pelos trabalhos deste subitem são referentes a bilhetagem eletrônica do transporte público, GPS para rastreamento de veículos ou indivíduos, a análise de dados locais gerados por usuários de sites e o uso da estrutura da telefonia celular no rastreamento de dispositivos. Há de se destacar que uma das principais vantagens dos métodos alternativos diante dos métodos tradicionais é que aquele não depende do respondente (participação ativa) para registros locais, as localizações são armazenadas automaticamente, por este motivo, é usual também denominar estas tecnologias de métodos passivos de coleta de dados em transporte.

Noulas et al. (2011) utilizaram check-ins do Foursquare, rede social em que o elemento principal é a localização do indivíduo que faz uso da rede, em resumo, os check-ins possibilitam descobrir se o indivíduo está no trabalho, estudo ou lazer. A

análise destes dados permitiram visualizar que nos finais de semana a categoria referente a escritórios deixa o topo das atividades mais desenvolvidas, sendo substituída por atividades relacionadas a lazer como bares e restaurantes. Long et al. (2012) utilizaram dados similares e obtiveram conclusões em consonância com Noulas et al. (2011), visto que concluíram que os check-ins dos finais de semana são majoritariamente referentes a entretenimento e recreação.

Zhong et al. (2015) utilizaram dados de bilhetagem eletrônica (*Smart Card*) do transporte público de Cingapura no período referente ao mês de abril de 2014. Os autores concluíram que, para a cidade analisada, há variabilidade no padrão de mobilidade a nível de indivíduo, entretanto, a estrutura geral dos deslocamentos urbanos permanecem quase a mesma. Depreende-se dessa conclusão que a nível agregado a variabilidade é substancialmente menor em comparação com padrões individuais.

Yu e He (2017) também fizeram uso de dados de *Smart Card* para análise da demanda por ônibus e traçaram conclusões relativas a como as condições do tempo interferem na mobilidade de uma região. Os autores descobriram que chuvas reduzem a demanda, mas não o padrão de distribuição geográfica dos deslocamentos, isto é, as chuvas apenas afetam na intensidade de demanda de viagens.

Elango et al. (2007) analisaram 1,5 milhões de viagens de um total de 480 veículos no período de três anos, os veículos foram equipados com aparelhos GPS. Eles descobriram que a variabilidade é semelhante entre segunda e quinta, havendo um aumento nas sextas-feiras. Entende-se então que as viagens, considerando segunda a quinta, tem grau consistente de repetição. Os autores ainda destacam que o dia de maior grau de variabilidade é sábado, o que pode ser explicado pelo fato deste dia ser frequentemente o mais atípico em diversas culturas, haja vista que existe uma tendência de que obrigações e atividades de lazer tenham maior equidade de tempo, em comparação a outros dias da semana.

Fazendo uso de GPS Stopher e Zhong (2011) coletaram dados de 50 indivíduos em um período de 28 dias. Os dados possibilitaram analisar não apenas o propósito da viagem, mas também atributos como distância percorrida, tempo de viagem e tempo de realização do propósito. Não observaram-se padrões repetitivos. Nos propósitos referentes a trabalho e estudo há alto grau de repetitividade, todavia, mesmo nessas viagens é baixo repetição de seus atributos, tais como o tempo de realização do propósito. Tal resultado chama atenção à assertiva de Schlich e Axhausen (2003) que

afirmam que a similaridade tende a declinar quanto maior for o grau de desagregação e complexidade dos dados obtidos.

Xianyu et al. (2015) utilizaram dados de um projeto de pesquisa que buscou correlacionar o uso de *smartphones* com o padrão de viagem humano, 46 respondentes participaram da pesquisa, sendo que 716 viagens foram reportadas. Os resultados indicam que dias da semana tem significativo efeito na variabilidade do padrão de mobilidade, com dias úteis tendo maior similaridade. Em média, cada respondente efetuou 2,6 viagens nos dias de semana e 2,3 viagens nos finais de semana. Os autores ainda destacam que na localidade de estudo (Xangai, China) observaram-se que padrões de quinta-feira são semelhantes a de fins de semana, com menor taxa de atividades relacionadas ao trabalho e menor mobilidade.

Na análise feita por Xianyu et al. (2015) ainda foi encontrado que os padrões dos dias úteis são significativamente diferentes dos padrões de finais de semana e a diferença entre padrões de dia útil e domingo é maior que a diferença entre dia útil e sábado.

Allstrom et al (2016) desenvolveram o sistema denominado MEILI que consiste em um aplicativo para *smartphone* para capturar movimentos de forma automática e uma plataforma web que dá possibilidade aos voluntários de reportar informações mais detalhadas acerca de suas viagens. Foram registradas 2142 viagens. Não foi encontrado um padrão significativo, os padrões variam de dia para dia, por exemplo, viagens a compras são feitas majoritariamente aos sábados, enquanto que atividades sociais são realizadas no domingo. Esportes e hobbies são realizados na quarta e sábado para a localidade de estudo (Estocolmo, Suécia).

Motahari et al. (2012) utilizaram a cobertura da telefonia móvel para rastrear 180 dispositivos celulares, os autores descobriram que pessoas se movem por maiores distâncias, escolhem lugares mais imprevisíveis para visitar e tem uma distribuição de probabilidade espacial mais dispersa nos fins de semana em comparação com os dias úteis. Sagl et al. (2012) também fazendo uso dos dados de telefonia móvel encontrou que há máxima mobilidade nas terças e quartas, sendo que nos outros dias úteis o grau de mobilidade se reduz. Domingo é o dia de menor intensidade de mobilidade, resultado em consonância com diversos estudos supracitados.

As Tabelas 2 e 3 elenca a bibliografia abordada neste trabalho. Percebe-se a grande dificuldade de se obter consenso acerca do tema variabilidade em transportes, talvez nem mesmo haja a possibilidade de que tal consenso seja atingido, levando em

conta o vasto número de características de indivíduos, famílias, culturas, sistemas de transportes e até mesmo as cidades como um todo. Esta dissertação é um esforço no caminho de propor tanto um conhecimento acerca de um dado ainda não utilizado na área de transportes, bem como estudar a variabilidade de um local distinto daqueles já analisados pela literatura.

Os dados utilizados neste trabalho tem características que se assemelham aos dados obtidos por métodos alternativos, com algumas particularidades. O próximo item é dedicada à apresentação destes dados.

Tabela 3 - Estudo da variabilidade com dados obtidos por métodos alternativos

Ano	Autor	Cidade ou País	Dado utilizado	Amostra	Período
2007	Elango et al.	<i>Atlanta</i>	GPS	480 veículos	3 anos
2011	Noulas et al.	N.D.	Foursquare	700.000 usuários	100 dias
2011	Stopher e Zhong	<i>Adelaide</i>	GPS	50 hab	28 dias
2012	Long et al.	Pittsburg	Foursquare	32.000 usuários	1 mês
2012	Motahari et al.	São Francisco	Antenas de telefonia	180 celulares	1 mês
2012	Sagl et al.	Norte da Itália	Antenas de telefonia	88 milhões de registros	2 meses
2015	Xianyu et al.	Xangai	App smartphone	46 hab	1 semana
2015	Zhong et al.	Cingapura	Bilhetagem eletrônica	5.000.000 transações/dia	7 dias
2016	Allstrom et al.	Estocolmo	App smartphone	2142 viagens	N.D.
2017	Yu e He	Guangzhou	Bilhetagem eletrônica	5.000.000 transações/dia	8 semanas
N.D. = Dado não disponível na bibliografia consultada					

Tabela 4 - Estudo da variabilidade com dados obtidos por métodos tradicionais

Ano	Autor	Cidade ou País	Dado utilizado	Amostra	Período de coleta
1981	Hanson e Huff	<i>Uppsala</i>	<i>Uppsala Household Travel Survey</i>	149 hab	35 dias
1986	Huff e Hanson	<i>Uppsala</i>	<i>Uppsala Household Travel Survey</i>	149 hab	35 dias
1988	Jones e Clarke	Reino Unido e Adelaide	<i>UK National Travel Survey e the Adelaide Travel Demand</i>	N.D	N.D
1988	Huff e Hanson	<i>Uppsala</i>	<i>Uppsala Household Travel Survey</i>	149 hab	35 dias
1986	Pas e Koppelman	<i>Reading</i>	<i>Reading Activity Diary Survey</i>	112 hab	3 meses
1988	Pas	<i>Reading</i>	<i>Reading Activity Diary Survey</i>	112 hab	3 meses
1995	Pas e Sundar	Seattle	Dados da Prefeitura de Seattle	150 domicílios	2 meses
2003	Schlich e Axhausen	Halle	Mobidrive	361 hab	6 semanas
2005	Susilo e Kitamura	Halle	Mobidrive	361 hab	6 semanas
2009	Tarigan e Kitamura	Halle	Mobidrive	361 hab	6 semanas
2010	Ahas et al.	Tallin	Questionário	277 hab	8 dias
2010	Kang e Scott	<i>Toronto</i>	<i>Toronto Travel-Activity Panel Survey</i>	192 hab	N.D.
2015	Dharmowijoyo et al.	Bandung	Pesquisa Própria	732 hab	21 dias
2017	Dharmowijoyo et al.	Bandung	Pesquisa Própria	732 hab	21 dias
2015	Dharmowijoyo et al.	<i>Jacarta</i>	<i>SIMTRAMP Household Travel Survey</i>	3846 hab	3 semanas ¹⁰
2016	Yang et al.	<i>Pequim</i>	<i>Beijing Comprehensive Travel Survey e Beijing Holiday Travel Survey</i>	10.650 e 5.690 hab	N.D
2017	Raux et al.	Gante	Dados próprios	717 hab	1 semana
N.D. = Dado não disponível na bibliografia consultada					

3. MOZILLA LOCATION SERVICE

O *Mozilla Location Service* (MLS) é um serviço de localização aberto, alimentado por colaboradores, baseado em antenas telefônicas e pontos de Wi-Fi. Os serviços de localização tem grande relevância e utilidade atualmente, uma vez que eles possibilitam o desenvolvimento de aplicativos que tem a capacidade de estimar o tráfego em um via, calcular o tempo de viagem, encontrar estabelecimentos próximos, entre outros serviços.

Por serem, em maioria, baseados em GPS, em diversas situações os serviços podem apresentar problemas no funcionamento, como por exemplo em lugares fechados ou em dias nublados. No intuito de contornar tais problemas foram criados novos meios de obtenção de dados de localização, tais como os meios utilizados pelo MLS. O Mozilla Location Service (MLS) utiliza pontos de Wi-Fi e antenas telefônicas para estimar a localização de um indivíduo.

A Mozilla é uma comunidade mundial de tecnologia que tem como anseio a garantia de que a internet seja um recurso público global, aberto e acessível a todos. Este ideal comunitário e colaborativo norteia diversos projetos da comunidade, tal como o Mozilla Firefox, navegador de código aberto que agrega feedbacks de desenvolvedores de todo o mundo e o Mozilla Location Service que também é sustentado por meio de colaborações.

As colaborações dos indivíduos são essenciais ao funcionamento do serviço de localização do Mozilla, pois ele funciona triangulando os sinais de vários pontos transmissores de sinalização, no caso torres de telefonia e pontos Wi-Fi, tais pontos devem ter localização conhecida. Para realizar estimativas locais com precisão, o provedor de serviços de localização necessita ter uma quantidade muito grande de dados relativos às localizações dos pontos Wi-Fi e antenas telefônicas do mundo (ALEKSANDERSEN, 2017). Neste momento, no que toca a coleta de dados, o trabalho colaborativo é importante, uma vez que, por meio do app Mozilla Stumbler, os colaboradores podem fazer a coleta de dados de localização de antenas e pontos Wi-Fi, a fim de possibilitar o funcionamento do serviço.

3.1. App Mozilla Stumbler

O Mozilla Stumbler é um aplicativo utilizado pelos colaboradores para coleta de dados de localizações de pontos Wi-Fi e antenas telefônicas, tal app tem a capacidade de estimar a localização destes pontos. O app, tampouco do Mozilla

Location Service, deixa claro a maneira com que é feita a estimativa da localização da antena, supõe-se que a intensidade do sinal é o principal elemento utilizado. Qualquer indivíduo que tenha um smartphone pode ser um colaborador, os colaboradores do MLS fazem a coleta dos dados de localização necessários ao aprimoramento do serviço.

O conjunto de dados coletados pelos colaboradores é disponibilizado em regime diário. Os dados locais são referentes a uma estimativa da localização da antena, que corresponde a um caminho próximo ao que o colaborador percorreu. Tais dados foram utilizados neste trabalho para estudo da variabilidade diária das viagens. Por dia, um total de 500Mb de um arquivo de texto (.csv) é gerado pelo serviço (disponível em: <https://location.services.mozilla.com/downloads>), contém, dentre outros dados, as localizações das antenas que se aproxima do percurso de um colaborador anônimo. Na Tabela 5, é realizada a descrição deste dados.

Os primeiros cinco campos da Tabela 5 (rádio a célula) identificam uma única rede celular lógica, isto é, tecnologia oferecida, operadora de telefonia e identificadores da célula (MOZILLA, 2013). Os campos restantes contêm informações relativas a esta rede.

Todas as informações locais são compiladas para criar uma espécie de mapa global no qual qualquer dispositivo que acessar uma das redes identificadas poderá obter a sua geolocalização aproximada (ALECRIM, 2013). Como já mencionado, a geolocalização é resultante de um processo de triangulação entre diversos sinais de Wi-Fi e antenas telefônicas.

No item seguinte é exemplificado o mecanismo de registro e atualização dos dados locais do Mozilla Location Service.

Tabela 5 - Dados fornecidos pelo MLS

Dado	O que eles representam
Radio	Tipo de rede. Pode ser 2G (GSM), 3G (WCDMA) ou 4G (LTE). Cada dispositivo irá coletar informações relacionadas à sua rede apenas, de acordo com a tecnologia disponível no dispositivo.
MCC	<i>Mobile Country Code</i> . É o código de três dígitos do país, padronizador pela IMSI (<i>International Mobile Subscriber Identity</i>), o código do Brasil é 724.
NET	Referente ao <i>Mobile Network Code</i> . É o código referente à operadora que o celular faz uso. Em celulares que são atendidas por duas ou mais operadoras, apenas o chip da operadora principal (geralmente ocupa o <i>slot</i> 1) está apto a coletar informações locais.
Area	Relativo a um parâmetro de telefonia denominado Location Area. É uma área na qual estão localizadas várias antenas.
Cell	Esta coluna se refere ao número identificador da célula.
Lon	Longitude aproximada do transmissor de sinal da célula
Lat	Latitude aproximada do transmissor de sinal da célula
Range	Estimativa de cobertura da célula, isto é, um raio no entorno da posição estimada que compreenderia toda a célula.
Samples	Número total de observações realizadas para calcular a posição e cobertura estimadas.
Created	Momento em que o registro da célula foi criado pela primeira vez
Updated	Momento em que o registro da célula foi atualizado

3.2. Registro e atualização dos dados obtidos pelos colaboradores

Neste subitem foi feita uma explanação do mecanismo de registro e atualização de dados do Mozilla Location Service. As afirmações feitas aqui são fruto de testes e conclusões a partir destes testes. No intuito de consolidar tais afirmações foram realizadas tentativas de contato com a Mozilla, entretanto, não houve sucesso. Na Figura 4 é retratada a situação na qual uma antena ainda não tenha sido ‘observada’ por um colaborador do MLS. Cabe salientar que a colaboração é realizada por meio do software Mozilla Stumbler.

Os cinco primeiros campos da tabela contida na Figura 4 são referentes às características da rede celular e, por isso, são fixas, não se modificam com colaboradores diferentes. Já os outros seis campos mudam de acordo com a quantidade de vezes com que a antena teve sua localização estimada. Os campos Lon e Lat são referentes à estimativa da latitude e longitude da antena telefônica que teve informações coletadas pelo app Mozilla Stumbler. Como a situação da Figura 4 é a primeira estimativa da latitude e longitude da antena observada, então, espera-se que tal estimativa seja distante da posição real e a medida que as observações forem feitas a estimativa fica mais próxima da localização real da antena.

O campo Range é referente ao raio de cobertura da antena observada, tais dados não foram utilizados neste estudo, entretanto, supõe-se que estes registros também tenham o mesmo comportamento que os campos Lat e Lon, isto é, vão ficando mais precisos a medida que as observações de uma mesma antena aumentam. O campo Samples é referente ao número de vezes em que uma antena foi observada. Na Figura 4, uma vez que se refere a primeira observação da antena, o campo Samples é 1, na segunda observação será 2, e assim por diante.

O campo Created é referente à data e hora com que a primeira observação foi feita, ou seja, a data e hora com que o registro da antena foi criado. Na Figura 4, o campo Created está no formato de hora (hh:mm:ss) a título de exemplo, no arquivo .csv do MLS tal campo está em formato UNIX que, após conversão, resulta na informação de data e hora. O campo Updated se refere à data e hora com que a última observação da antena foi realizada, isto é, a data e hora em que foi feita a atualização da localização da torre de telefonia observada. Cabe salientar que a localização da torre corresponde a localização aproximada do colaborador, e mesmo havendo erro considerável na estimativa da posição do colaborador, para uma área extensa, a precisão utilizada é suficiente. Para os objetivos deste trabalho, o campo Updated tem grande importância, visto que possibilita extrair a data de upload dos dados, que será considerado como o dia em que a viagem foi efetuada, bem como possibilita extrair um colaborador, conforme será explanado no item seguinte.

A Figura 4 ainda representa o erro de estimativa da posição da antena, tal erro foi denominado 'e'.

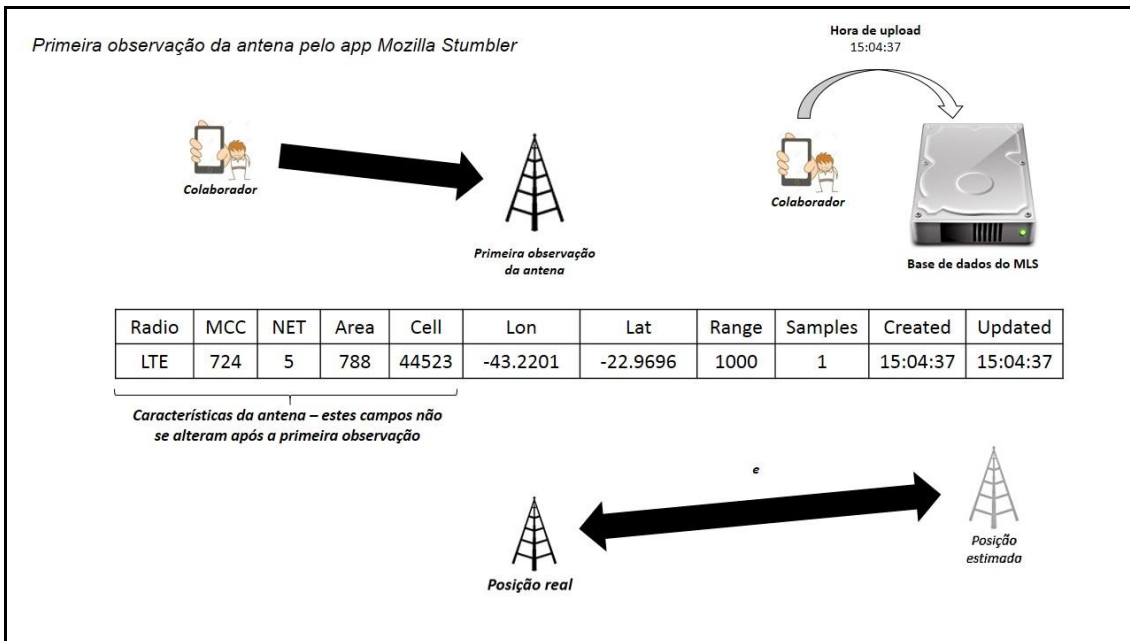


Figura 4 - Primeira estimativa de localização da antena

A Figura 5 é referente à décima observação da antena pelo app Mozilla Stumbler. Conforme mencionado anteriormente, os cinco primeiros campos não se alteram. Os campos Lat e Lon se modificam a cada nova detecção por parte dos colaboradores, espera-se que a cada detecção a posição estimada fique mais próxima da posição real da antena, uma vez que o objetivo do trabalho é o de proporcionar a melhor estimativa possível para a localização de uma antena.

Na representação do distância entre a localização real e a localização estimada percebe-se que o erro ‘ e ’ foi subtraído por $f(S)$, que é uma função que representa a influência do tamanho da amostra (Samples - S) na aproximação das localizações real e estimada.

O campo Created, conforme disposto anteriormente, é referente à data de criação do registro da antena, portanto, após a primeira observação este registro também passa a ser fixo. Observa-se, na Figura 5 que o campo Created é o mesmo da Figura 4. O campo Updated, referente à data e hora de atualização, modifica-se a cada upload de um dado colaborador, nota-se que o horário de upload coincide com o registro no campo Updated. O campo Samples, uma vez que se trata da décima observação, tem o número 10 como registro.

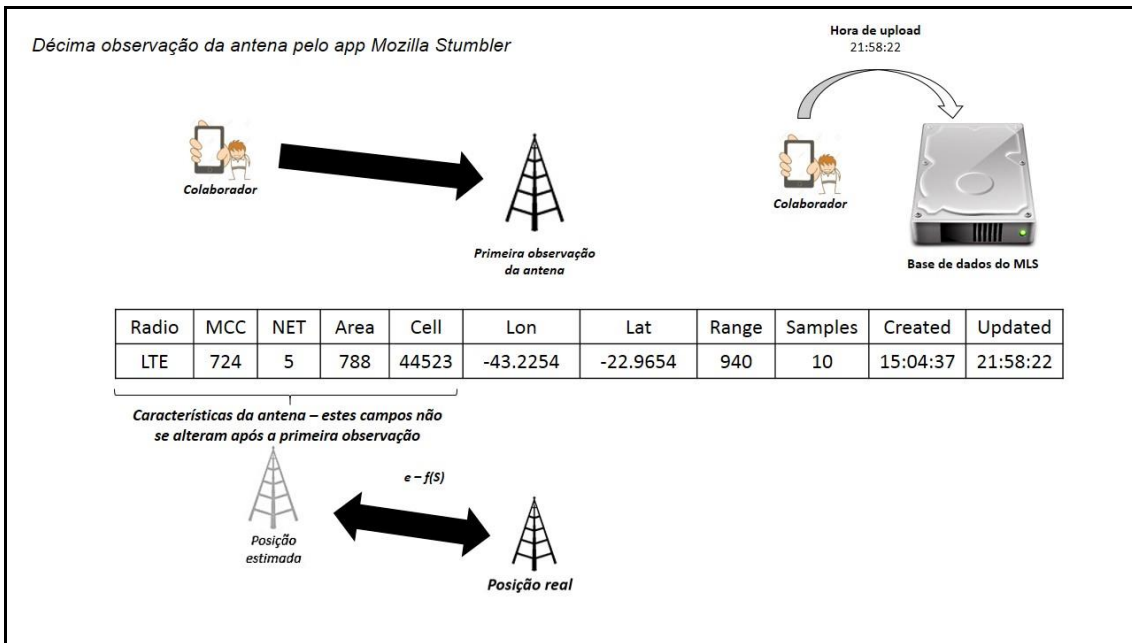


Figura 5 - Décima estimativa de localização da antena

Na Figura 6, esta retratada a situação em que a mesma antena das situações anteriores tem dados de localização coletados pela centésima vez. Conforme já descrito os primeiros cinco campos são imutáveis, além do campo Created que após a primeira coleta também não se modifica.

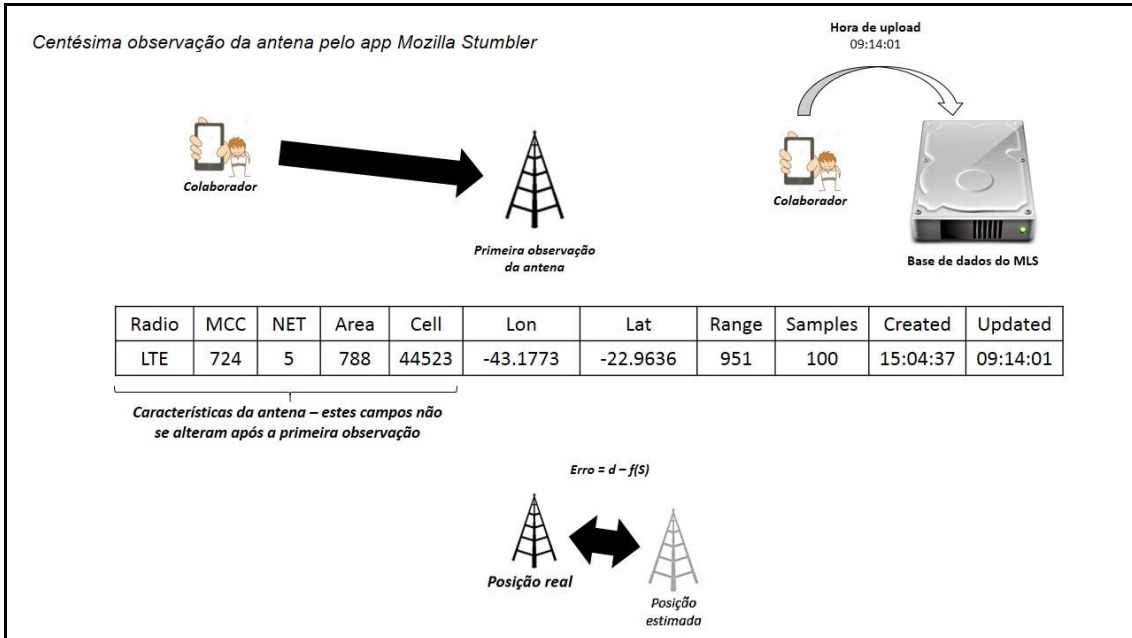


Figura 6 - Centésima estimativa de localização da antena

Os campos referentes às coordenadas geográficas se modificam e se aproximam cada vez mais da posição real da antena telefônica. Neste ponto o objetivo do trabalho colaborativo por meio do Mozilla Stumbler alcança êxito, uma vez que com a localização da antena próxima da realidade, as estimativas de localização por

triangulação (Figura 7) do Mozilla Location Service ficarão mais precisas. Isto se deve ao fato de que a triangulação apenas é possível devido ao conhecimento das localizações das antenas telefônicas. Como explicado por Aleksandersen (2016), os dispositivos podem ter suas localizações aproximadas pela comparação entre intensidades de sinais de diversos transmissores com localizações conhecidas. A localização é disponibilizada pelo ponto de interseção entre todos os sinais.

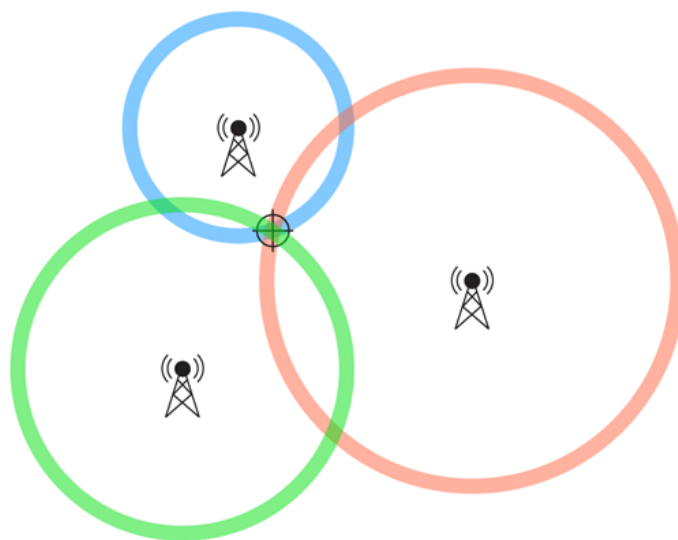


Figura 7 - Triangulação de antenas

Neste item buscou-se explicar sobre o funcionamento do Mozilla Stumbler, bem como seu papel para a obtenção de dados de antenas telefônicas do Mozilla Location Service. No item seguinte, foi feita a caracterização dos dados no intuito de direcioná-los aos estudos em transportes.

3.3. Caracterização dos dados fornecidos pelo MLS

Este subitem pretende explicar como são gerados os dados locais fornecidos pelo MLS, desde a coleta pelo colaborador, até a disponibilização diária do arquivo texto. Além disso, busca fornecer conhecimentos acerca da utilidade dos dados para estudos em transportes. No dia 19 de junho de 2017 foi realizado um teste com o aplicativo Mozilla Stumbler a fim de verificar de que maneira são gerados os dados locais do MLS. Os percursos de ida e volta estão representados nas Figuras 8 e 9. Cabe salientar que a coleta foi realizada tanto no sentido Centro, quanto no sentido Ilha do Fundão.



Figura 8 - Trajeto de ida

Fonte: Adaptado do app Moovit

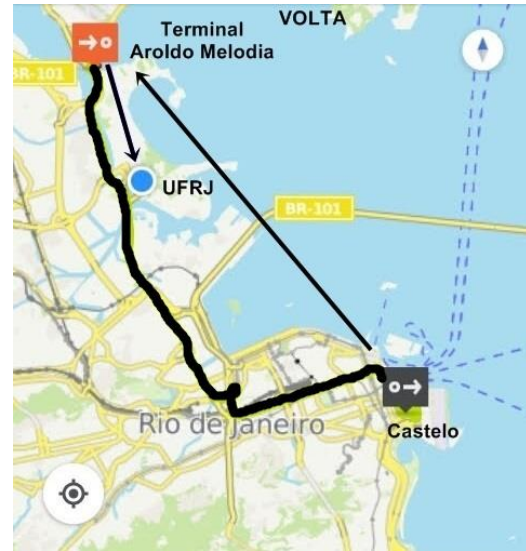


Figura 9 - Trajeto de volta

Fonte: Adaptado do app Moovit

Para que a coleta seja iniciada, o colaborador deve abrir o app Mozilla Stumbler. Com o Mozilla Stumbler aberto, basta iniciar a viagem para que o app colete informações locais de forma automática. Durante o percurso realizado no teste, o aplicativo Mozilla Stumbler permaneceu ativo, ou seja, coletando dados de pontos Wi-Fi e antenas telefônicas de seu alcance. A Figura 10 é referente a uma captura de tela do app, em um determinado momento da viagem. A coleta estava sendo realizada nas proximidades da Igreja da Candelária, conforme indicações.



Figura 10 - Screenshot no momento do teste

Ainda na Figura 10, a área predominantemente azul é referente ao local onde ao menos uma vez um colaborador coletou dados de localização para a base do MLS. Os círculos verdes são referentes a pontos onde foi feita a coleta de dados de Wi-Fi e antenas telefônicas, já os quadrados azuis são relativos a coleta de localizações de antenas telefônicas, somente. Cada símbolo está relacionado a um registro que posteriormente será carregado à base de dados disponibilizada publicamente. Esta base de dados é composta por antenas de todas as partes de mundo, conforme pode ser observado na Figura 11 referente aos pontos já coletados pelos colaboradores do MLS.

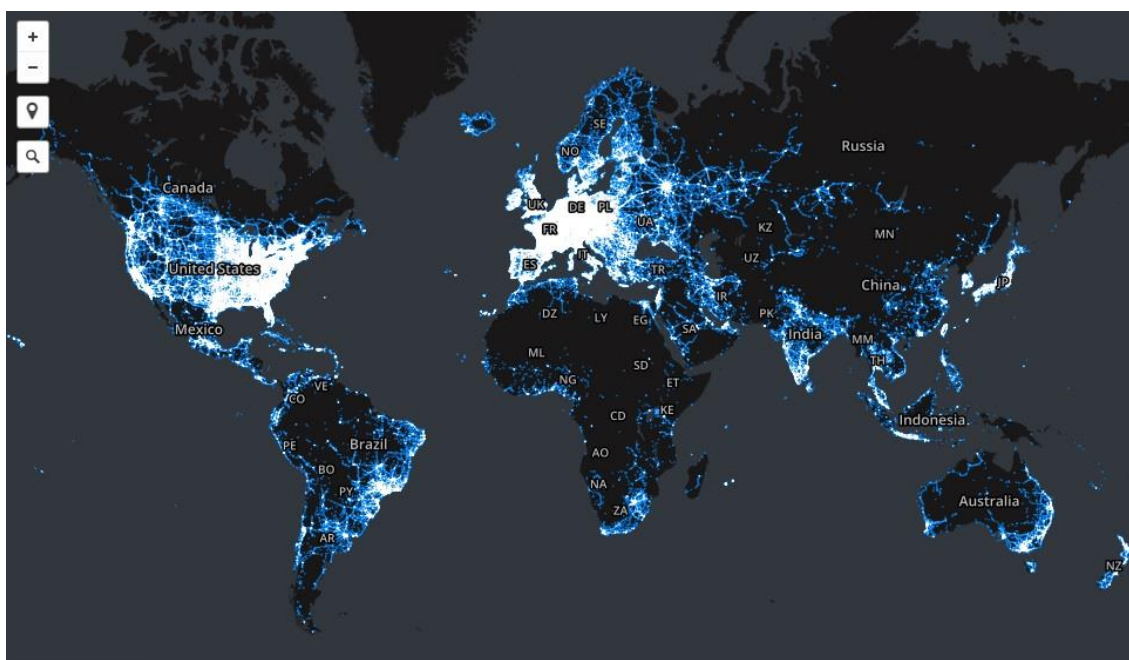


Figura 11 - Pontos coletados pelo colaboradores

Na Figura 11, as áreas mais claras representam áreas com maior número de detecções e, conseqüentemente, de colaboradores. No mundo se destacam os Estados Unidos e países da Europa no que tange à adesão ao projeto Mozilla Location Service, já no Brasil a região Sudeste ganha tal destaque.

As áreas mais claras tem maior número de colaboradores e maior probabilidade de antenas serem captadas. Entretanto, o fato de uma antena telefônica já ter sido captada por algum colaborador não significa que não será por outro colaborador também, pelo contrário, quanto mais vezes uma antena ou um ponto Wi-Fi for ‘visualizado’ pelo app, mais precisa será a localização estimada. Este processo que se baseia na iteração dos dados coletados pelos colaboradores foi explanada no item anterior e resumido na Figura 12.

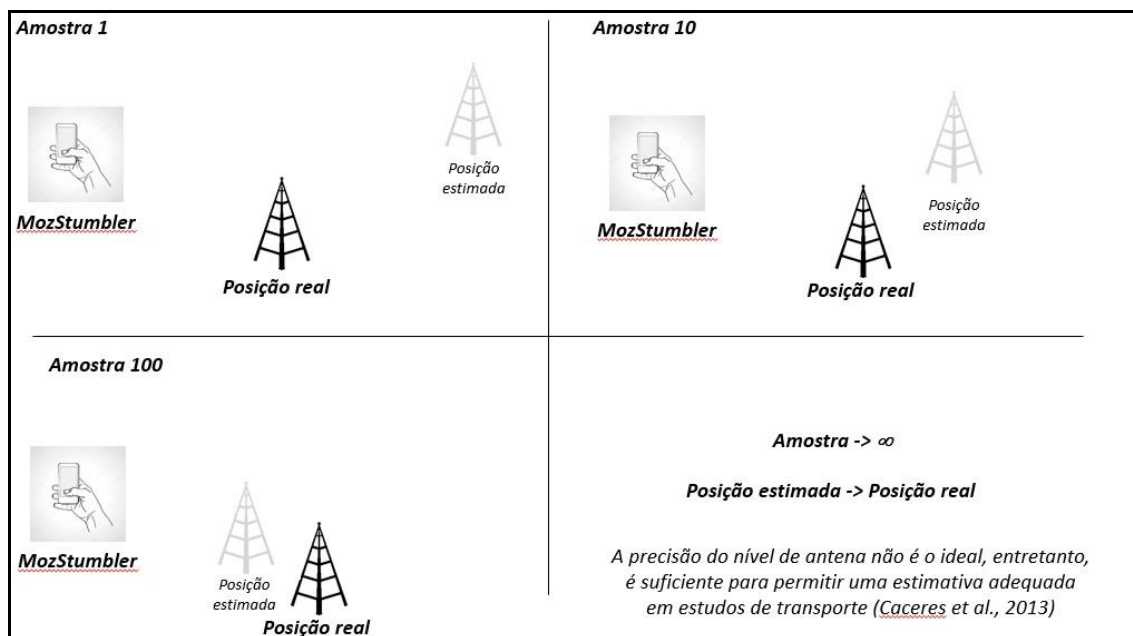


Figura 12 - Processo de estimação da localização da antena

Cabe salientar aqui uma situação: dois colaboradores diferentes coletam dados referentes a mesma antena e no mesmo dia. Pode-se pensar que apenas o último colaborador será visualizado. Entretanto, por meio da análise de diversas bases de dados fornecidas pelo MLS, notou-se que algumas antenas se repetiam diversas vezes em um mesmo dia. Concluiu-se, portanto, que as antenas são registradas na base de dados diária, independentemente de quantas vezes são ‘visualizadas’ pelo app Mozilla Stumbler em um mesmo dia.

O aplicativo Mozilla Stumbler permanece ativado quando o colaborador está em movimento, entretanto, o app entra em modo de espera quando o colaborador está parado. Isto quer dizer que o simples fato de abrir o app não é suficiente para gerar dados locais à base de dados MLS, o colaborador deve realizar uma viagem qualquer para que haja a geração dos dados. Entende-se, então, que **os dados fornecidos pelo MLS são estritamente vinculados a uma viagem**, uma vez que colaboradores que não estão em movimento não geram dados.

Após o deslocamento e coleta de dados pelo Mozilla Stumbler deve ser realizado o *upload* dos dados coletados na viagem. Uma característica importante é que o default do aplicativo, isto é, sua configuração padrão inicial é de fazer o upload apenas em redes Wi-Fi, ou seja, os dados não são enviados para a base MLS automaticamente, mesmo que o colaborador tenha rede de dados disponível em seu celular para realizar o upload.

Esta característica acarreta que o horário disposto na coluna *'updated'* que, conforme Tabela 5, é o horário de atualização da localização da antena, não corresponde ao horário de realização da viagem, mas sim o horário em que foi realizado o upload das informações coletadas na viagem para a base de dados do Mozilla Location Service. Então, não há a possibilidade de saber se o colaborador realizou a viagem em determinado horário, mas, devido ao fato de os dados serem disponibilizados diariamente, adota-se que a viagem foi realizada no dia de disponibilização dos dados.

A base de dados referente às antenas telefônicas sempre é disponibilizada ao final do dia, tal base contém as localizações atualizadas de todas as antenas que foram detectadas pelo aplicativo Mozilla Stumbler. No intuito de gerar conhecimentos acerca destes dados baixou-se a base de dados referente ao dia do teste na página de downloads do site do MLS.

Após o download da base de dados foi realizado o procedimento de tratamento das informações contidas no arquivo .csv. Tal procedimento será detalhado no capítulo seguinte, referente ao método proposto. Por meio da visualização dos dados, foi possível notar uma outra característica importante dos dados captados pelo Mozilla Stumbler. O horário de upload foi 13:47:20, verificou-se que este horário se repetia por várias vezes na coluna *'updated'* do banco de dados. Ao plotar tais os registros relativos a este horário no mapa do Rio de Janeiro, percebeu-se que se tratava do caminho onde teste foi realizado (Figura 13), ou seja, os dados referentes ao teste realizado foram carregados em um mesmo instante, sendo atribuídos o mesmo valor ao campo *'updated'*.

Esta característica é uma das mais importantes acerca dos dados estudados, pois possibilita inferir que registros com instantes iguais são provenientes de um mesmo colaborador, ou seja, a coluna *'updated'* contém um possível identificador para o colaborador, ainda que este identificador seja válido apenas para o dia analisado, haja vista que em outro dia o colaborador pode realizar o upload em outro horário.

Tal característica é importante pois minimiza a deficiência intrínseca aos dados estudados, que é a não disponibilidade de um id por colaborador. Sabendo que um colaborador tem o mesmo horário de upload é possível inferir o número de colaboradores, bem como o número de viagens realizadas por todos os colaboradores. Entretanto, cabe salientar que podem ocorrer algumas situações que não são possíveis de serem observadas por meio dos dados fornecidos pelo MLS.



Figura 13 - Plotagem dos pontos fornecidos pelo MLS

São duas as situações elencadas aqui. A primeira é a quantidade de pontos Wi-Fi que um colaborador tem acesso durante o dia. Por exemplo, um colaborador pode iniciar uma viagem de sua casa até o trabalho e quando chegar ao local de trabalho realizar o upload dos dados coletados no deslocamento casa-trabalho. O mesmo colaborador pode coletar dados no trajeto de volta, isto é, trajeto trabalho-casa, e realizar o upload dos dados em casa também. Neste caso o horário de upload seriam dois, entretanto, o colaborador seria o mesmo. Nesta situação, considerar o campo 'updated' como um identificador acarretaria em duplicar um determinado colaborador. A segunda situação se refere à assiduidade do colaborador. Um colaborador pode carregar os dados no mesmo dia da viagem realizada, mas também o mesmo colaborador pode não ter acesso a uma rede Wi-Fi durante o dia. Neste caso, os dados apenas serão carregados quando o indivíduo tiver acesso a uma rede, o que levaria à impossibilidade de afirmar a data de coleta pelo colaborador. Este trabalho fez considerações a fim de tornar possível o estudo de tais dados, por meio da suposição de que o indivíduo é assíduo e também carrega os dados apenas uma vez por dia.

Retomando a análise da Figura 13, uma observação a ser feita é que os pontos de coleta que aparecem no app, os círculos verdes e quadrados azuis (Figura 10), não necessariamente são os mesmos pontos que são disponibilizados no final do dia. Isso ocorre em consequência de que, quando ativado, o app marca a coordenada geográfica onde o dispositivo está, bem como informações sobre o dado coletado em tal

coordenada, isto é, se é referente a antenas telefônicas (quadrados azuis), redes Wi-Fi's ou ambos os pontos (círculos verdes). Já o banco de dados fornece as informações atualizadas acerca das localizações das torres de telefonia, que comumente diferem das localizações GPS dos dispositivos.

Outra observação a se fazer é que as detecções na região da Ilha do Fundão são bem mais esparsas e em menor número do que no Centro. Isto se deve a dois fatores, o primeiro é o menor número de antenas na região da UFRJ em comparação com o Centro (Figura 14), e o segundo é que a velocidade média na Ilha do Fundão é bem mais alta, pois no Centro os problemas com congestionamentos são constantes. A baixa velocidade dá possibilidade ao aplicativo realizar a coleta de mais informações referentes a antenas do que em velocidades mais altas.

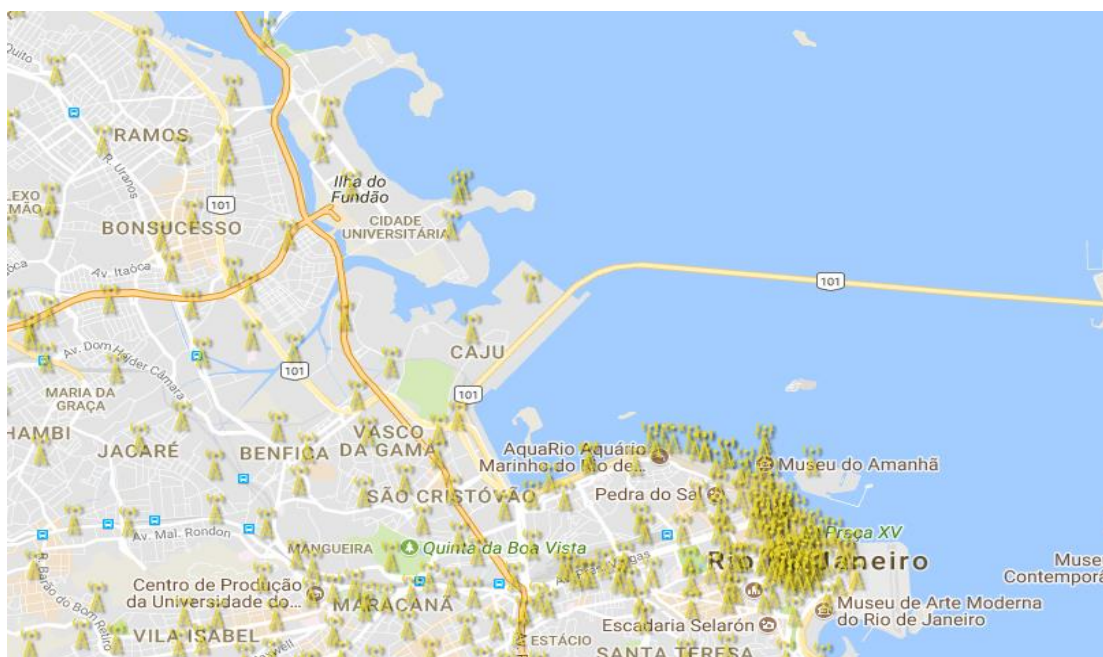


Figura 14 - Antenas da região que foi realizada a coleta

Fonte: Associação Brasileira de Telecomunicações (2017)

Ao comparar os pontos fornecidos pelo MLS (Figura 13) com os dados de localizações reais das antenas (Figura 14) são perceptíveis diversas imprecisões, isto se deve ao fato de que a localização das antenas é feita de maneira iterativa, isto é, aproxima-se da realidade a cada detecção. Por este motivo é recomendado aos colaboradores que utilizem o app de maneira frequente, pois mesmo os lugares que já tenham sido rastreados pelo Mozilla Stumbler tendem a melhorar a estimativa de posição a cada vez que o ponto é observado.

Percebe-se também que existem pontos muito próximos um ao outro (Figura 13), isto se deve à possibilidade de existirem várias células por antena telefônica.

Conforme explanado na revisão da literatura, as antenas podem ser omnidirecionais ou setorizadas, sendo que para cada tipo de antena há uma quantidade de células diferentes, para as antenas omnidirecionais é uma célula, já para as antenas setorizadas varia de duas a três células. O aplicativo, ao capturar os sinais de todas as células de uma antena, tende a gerar registros com localização muito próximas uma das outras, uma vez que as células pertencem à mesma antena. No intuito de simplificar as explicações optou-se por usar apenas a palavra antenas, no sentido que abrange tanto as células, quanto as antenas.

Além das imprecisões relatadas, os dados ainda estão sujeitos à questão relacionada ao congestionamento das antenas, isto é, em antenas sobrecarregadas os usuários são transferidos para uma rede mais distante, mas menos carregada. Devido a estas imprecisões, inerentes ao dado utilizado, recomenda-se trabalhar a nível de macrozonas, uma vez que os erros ainda que consideráveis em comparação com o GPS, tendem a estar na mesma macrozona por onde o colaborador MLS realmente trafegou. Portanto, no nível de macrozona os erros de localização tendem a ser diluídos ou ‘englobados’ em uma, o que minimiza seus efeitos nos cálculos finais.

Em resumo os dados fornecidos pelo MLS desde a sua coleta até a disponibilização segue a sequência disposta nos itens abaixo.

1. O indivíduo inicia a sua viagem, o aplicativo inicia a detecção;
 - a. Todas as antenas telefônicas ao alcance tem seus dados coletados
 - b. O app faz estima a distância do celular à antena
 - c. O app só funciona sob movimento do colaborador

2. Em um ponto Wi-Fi os dados são carregados, todos em um mesmo instante;
 - a. As localizações de antenas telefônicas estimadas pelo Mozilla Stumbler são mescladas à base de dados do Mozilla Location Service
 - b. Os campos **lat** e **lon** (coordenadas geográficas estimadas da antena telefônica) são atualizados por um método iterativo que leva em conta a localização anterior e a estimada por um novo dispositivo
 - c. O campo **updated** é atualizado com o horário de upload feito pelo colaborados

3. Os dados são disponibilizados maneira pública;
 - a. Dados referentes ao dia de coleta são fornecidos por volta da meia-noite
 - b. Contém os dados referentes às localizações atualizadas das antenas telefônicas

Os dados são disponibilizados diariamente no período noturno. A página de downloads do Mozilla Location Service contém o banco de dados de três dias, após três dias disponibilizados os arquivos são excluídos, o que impossibilita sua obtenção. Os arquivos são .csv, basicamente textual, de aproximadamente 500 Mb. Cada arquivo contém cerca de 22 milhões de registros, referentes às antenas já visualizadas pelo Mozilla Stumbler pelo mundo. Deve-se observar que o arquivo contém 22 milhões de registros de antenas, entre aquelas que foram captadas no dia ou anteriormente a data de disponibilização do arquivo, ou seja, nem todos os registros foram coletados ou atualizados na data em que foram fornecidos.

No próximo capítulo é feita a explanação do método proposto. Tomou-se como base a principal conclusão resultante do teste realizado, ou seja, a conclusão de que o colaborador pode ser identificado pelo horário de realização do upload.

4. MÉTODO PROPOSTO

Esta pesquisa objetiva obter o grau de variabilidade do índice de mobilidade para a cidade do Rio de Janeiro. Para isso, foram usados dados provenientes do *Mozilla Location Service* nos dias correspondentes à dia útil, final de semana e feriados. Pretende-se ainda fornecer conhecimentos acerca dos dados do MLS, que são dados públicos e ainda não utilizados para estudos em transportes. Por último, objetiva-se estudar a variação da mobilidade de uma cidade com características distintas daquelas cidades já analisadas por trabalhos anteriores.

Esta pesquisa, quanto à abordagem, é quantitativa, uma vez que os resultados podem ser quantificados e recorre-se à matemática para construção do conhecimento, descrição de causas, relações entre variáveis, etc. Quanto à natureza a pesquisa é aplicada, visto que objetiva gerar conhecimentos para aplicação prática na área de transportes.

Quanto aos objetivos, esta pesquisa busca descrever a variação do índice de mobilidade entre vários dias por meio de um estudo de caso na cidade do Rio de Janeiro. Quanto aos procedimentos, a pesquisa é um estudo de caso, visto o fato de se analisar um grupo (colaboradores do MLS), a fim de gerar conhecimentos acerca da variação do índice de mobilidade. Levando-se em conta que a cidade tem características distintas das demais estudadas na literatura, pode-se considerar a cidade como um grupo ao qual a análise será realizada.

Nos subitens seguintes foi realizada uma explanação acerca do método de tratamento e análise dos dados. Primeiramente, procurou-se elencar as considerações feitas por este trabalho no que se refere aos dados estudados, haja vista que diversas incertezas estão intrínsecas aos mesmos. Após as considerações acerca dos dados, foi apresentado o método de tratamento, desde a coleta pelos colaboradores, passando pela obtenção da base de dados diária, até a contagem do número de viagens por macrozona. Por último, apresentou-se uma proposta de análise estatística das variações possivelmente observadas no índice de mobilidade.

4.1. Considerações adotadas

Neste subcapítulo foi realizada a explanação das considerações que tornam possível o uso dos dados fornecidos pelo MLS para estudos de transportes. As considerações são duas:

1. O colaborador carrega os dados uma vez por dia;

2. Os usuários são frequentes/assíduos

A primeira premissa é feita devido a impossibilidade de saber se há duplicidade do colaborador, uma vez que não é fornecido um identificador para cada um. Tal consideração acarreta que cada registro no campo *updated* da base de dados é referente a um único dispositivo, o que possibilita o estudo do deslocamento para um dia. A segunda premissa garante que há assiduidade no uso do aplicativo, isto é, se o colaborador efetuar uma viagem, ele estará com o aplicativo Mozilla Stumbler ativado e, ao final da viagem realiza o carregamento dos dados. Tais premissas fragilizam as conclusões obtidas pelo trabalho, entretanto, foram necessárias à medida que se chegou ao entendimento da deficiência dos dados para estudos em transportes. O declínio das premissas está condicionado à estudos futuros relativos aos dados do Mozilla Location Service.

4.2. Tratamento dos dados

Neste subcapítulo foi realizada uma abordagem acerca do tratamento dos dados do Mozilla Location Service a fim de estudar a variação do índice de mobilidade para uma cidade qualquer. Este subcapítulo foi dividido em itens que dão uma visão detalhada desde a coleta dos dados pelos colaboradores, através do app Mozilla Stumbler, passando pela obtenção dos dados na base do MLS, até a filtragem e contagem dos dados a fim de chegar ao índice de mobilidade para uma determinada região.

4.2.1. Obtenção dos dados

Os dados do Mozilla Location Service são gerados por meio de colaboração de indivíduos que possuem e utilizam o aplicativo Mozilla Stumbler em seus deslocamentos diários. O Mozilla Stumbler coleta dados de pontos Wi-Fi e antenas telefônicas que estão ao alcance do dispositivo no trajeto desenvolvido pelo colaborador. Qualquer pessoa que tenha um smartphone pode colaborar com a base de dados MLS, bastando para isso ativar o aplicativo ao realizar viagens. O Mozilla Stumbler é vinculado ao movimento de seus colaboradores, só há funcionamento do aplicativo por meio de viagens, uma vez que o app entra em *stand by* quando o colaborador não está em movimento.

Ao final da viagem ou de um encadeamento de viagens, o colaborador realiza o upload dos arquivos de localização para a base de dados do Mozilla Location Service.

Tais registros atualizam diversos campos dentre os fornecidos pelo banco de dados MLS e dispostos na Tabela 5. Os registros atualizados são: Lon, Lat, Samples e Updated. Os registros não atualizados são relativos a características da rede, bem como identificadores da célula da antena telefônica, que são registros que não se modificam.

Os dados são disponibilizados diariamente no período noturno. A base de dados contém as atualizações diárias das localizações (lat e lon) das antenas telefônicas as quais uma rota pode ser extraída. A obtenção dos dados é feita por meio do download das bases de dados de localizações de antenas telefônicas no site do Mozilla Location Service.

Por se tratar de um trabalho acerca de variação do índice de mobilidade entre dias distintos, recomenda-se que sejam feitos downloads de, no mínimo, 15 dias, conforme recomendado por Huff e Hanson (1988). Tal medida deve ser seguida a fim de observar a variabilidade entre os dias da semana, bem como as variações no índice de mobilidade entre os dias úteis ou finais de semana.

4.2.2. Filtragem dos dados

Das 11 colunas disponibilizadas pelo banco de dados MLS, dispostas na Tabela 5, utilizam-se três, que são **lat** (latitude da antena), **lon** (longitude da antena) e **updated** (momento de upload da localização das antenas). Tais campos são justamente aqueles que são atualizados constantemente pelos colaboradores. Sendo que o campo ‘updated’ tem importância notável, visto que possibilita identificar um determinado colaborador.

Os dados utilizados permitem extrair uma rota similar à realizada pelo colaborador, por meio dos dados de latitude e longitude. Também torna possível identificar um colaborador por meio da repetição do horário de upload, o que indica que o carregamento foi realizado por uma mesma pessoa. As oito colunas restantes (radio, MCC, net, área, cell, range, samples, created) são relativas às configurações da rede de telefonia celular, tal como a operadora utilizada pelo colaborador e a tecnologia (2G, 3G ou 4G) do dispositivo utilizado pelo colaborador. Tais colunas não interferem na obtenção das detecções, portanto, foram descartadas.

Quanto às linhas (ou registros), são compostas por campos numéricos que representam as coordenadas geográficas das antenas do mundo todo. Por este motivo a quantidade de registros é tão alta, alcançando mais de 20 milhões por arquivo disponibilizado. Também é fornecido o horário de upload em formato UNIX, que se refere ao número de segundos passados desde a data de 1 de janeiro de 1970,

denominada *The Epoch*. Este formato para horário é incomum, entretanto, facilita o tratamento de dados em programas como o MySQL.

A filtragem foi realizada tanto em colunas quanto nas linhas. No que toca as colunas, foram descartadas as oito colunas supracitadas no intuito de reduzir a quantidade de dados e, conseqüentemente, aumentar a eficiência do software Access. Já a filtragem das linhas se fez necessária devido ao fato de que os registros são referentes às coordenadas do mundo inteiro, com isso, deve-se filtrar as coordenadas geográficas (lat e lon) referentes a área de estudo.

4.2.3. Atribuição de uma área às coordenadas geográficas

As coordenadas por si só não trazem conhecimentos interessantes acerca de deslocamentos entre áreas. É necessário atribuir a área onde cada coordenada geográfica está contida. Recomenda-se que a área escolhida seja semelhante às zonas ou macrozonas estabelecidas pela prefeitura da cidade de estudo. Após a atribuição das áreas será possível detectar de maneira mais fácil os deslocamentos diários de um dado colaborador. Por exemplo, suponha que um colaborador colete dados referentes a Área 1, Área 2 e Área 3, conforme representados na Figura 15.

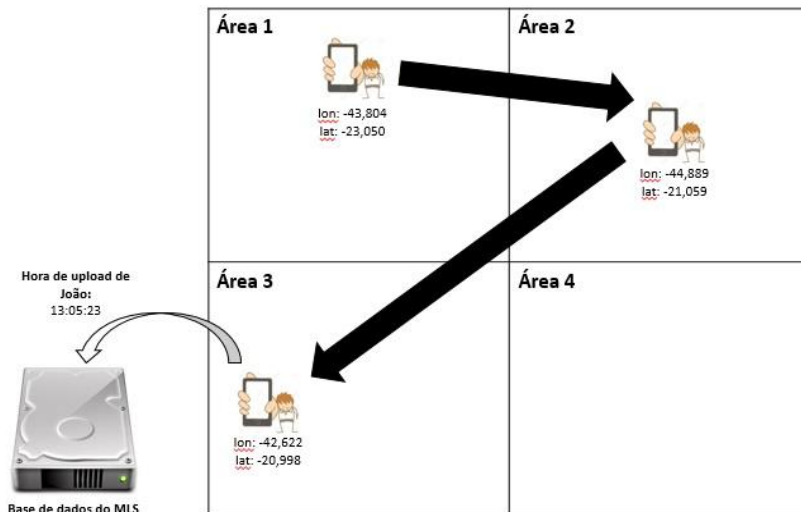


Figura 15 - Atribuição de área às coordenadas geográficas

Ao ter acesso a um ponto Wi-Fi na Área 3, os dados são carregados à base do MLS. Conforme explanado anteriormente, este colaborador pode ser identificado por meio do momento que o upload foi realizado, neste caso o carregamento ocorreu no horário 13:05:23. Com isso, as coordenadas do deslocamento disponibilizadas pela base de dados estarão juntas ao horário de realização de upload, que fica no campo 'updated', conforme exemplificado nas três primeiras linhas da Tabela 6.

Tabela 6 - Exemplo de atribuição Pt. 1

Lat	lon	Updated
-43,804	-23,050	13:05:23
-44,889	-21,059	13:05:23
-42,622	-20,998	13:05:23
-44,959	-21,460	22:00:55
-44,777	-21,000	22:00:55
-44,889	-21,002	22:00:55

Tabela 7 - Exemplo de atribuição Pt. 2

Updated	Área
13:05:23	1
13:05:23	2
13:05:23	3
22:00:55	2
22:00:55	2
22:00:55	2

A atribuição de áreas fará com que seja imediata a visualização da viagem realizada pelo colaborador (Tabela 7). Além disso, permite verificar se as viagens foram interzonais ou intrazonais, tal como o colaborador referente às 3 últimas linhas. Com isso, a atribuição de área tende a fornecer conhecimentos importantes acerca das viagens na área de estudo, pois facilita a verificação da quantidade de viagens por zona, que resulta no índice de mobilidade.

4.2.4. Agregação dos dados

O teste mostrou que uma viagem realizada por um dado colaborador pode ser detectada pelo horário de upload, que está armazenada no campo 'updated' da base de dados do Mozilla Location Service. Para cada viagem são criados vários registros compostos por diferentes lat e lon, que são referentes às localizações atualizadas das antenas telefônicas e que se assemelham ao trajeto realizado pelo colaborador.

Para cada registro, por haver coordenadas geográficas, pode ser atribuída uma determinada área, a qual fica a cargo do pesquisador definir, conforme explanado no subitem anterior. Tem-se, então, após a atribuição da área, vários registros com a mesma hora de upload, o que identifica o colaborador, bem como vários registros contendo a área onde o colaborador esteve, o que indica a rota realizada por tal colaborador.

Neste ponto cabe retomar o conceito de índice de mobilidade. Conforme Ferraz e Torres (2004) índice de mobilidade é a quantidade de viagens urbanas realizadas. Este trabalho fez uma adição ao conceito de índice de mobilidade formulado por estes autores, a fim de considerar a divisão por áreas realizada no tratamento de dados. Assim, índice de mobilidade é a quantidade de viagens urbanas realizadas por área.

Para entender como o agrupamento de registros fornecidos pelo MLS é útil ao cálculo do índice de mobilidade, deve-se entender que: os dados do Mozilla Location Service são gerados apenas sob viagens, haja vista que o app não coleta dados de um

colaborador inerte. Portanto, 1 registro de horário de upload (indica o colaborador) e 1 registro de área (indica provável rota) é capaz de indicar uma viagem realizada numa determinada área da cidade.

A Figura 16 ilustra um exemplo de como a agregação dos dados ajuda o processamento de dados da base MLS. O colaborador 1 realizou uma viagem entre as áreas 1, 2 e 3, gerou um total de 8 registros para a base de dados diária do MLS. O colaborador 2 realizou uma viagem e gerou 3 registros, referentes às áreas 1, 3 e 4, tais registros estão dispostos na Tabela 8. São vários os fatores que podem explicar a diferença entre o número de registros, tais como tamanho da viagem, velocidade da viagem, número de antenas no trajeto ou tempo de atividade do aplicativo Mozilla Stumbler. Entretanto, ambos os colaboradores tem algo em comum, eles realizaram uma viagem em determinadas zonas. Com isso, saber 1 registro referente ao colaborador e 1 registro referente a cada zona onde o colaborador trafegou, é o suficiente para concluir que o colaborador realizou uma viagem por determinada zona. O somatório das viagens realizadas por cada colaborador resulta no índice de mobilidade

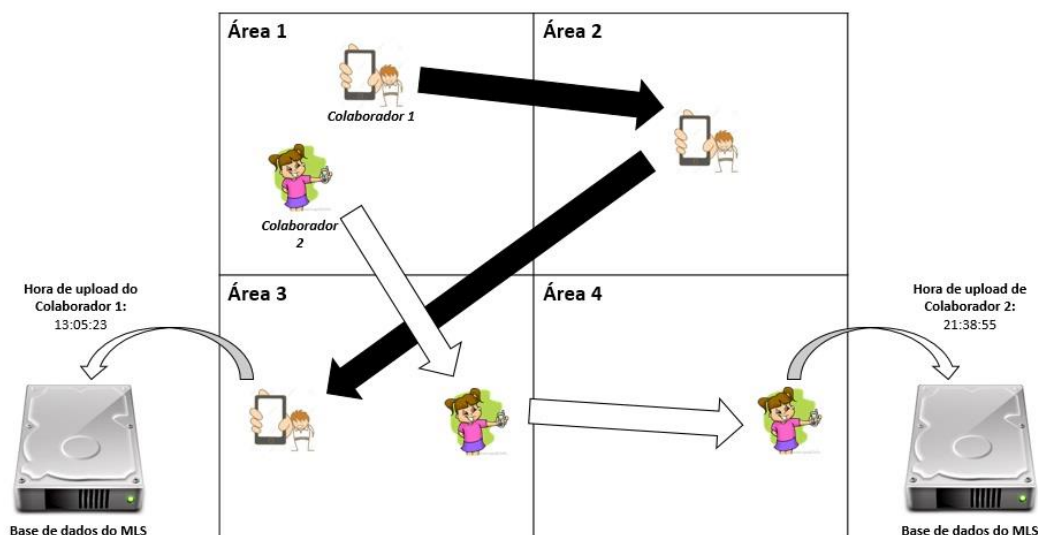


Figura 16 - Agregação dos dados MLS

Então, tem-se que os 8 registros gerados pelo Colaborador 1 podem ser agrupados na quantidade de zonas distintas que ele passou, o mesmo acontece com o Colaborador 2. Os registros resultantes desse agrupamento estão dispostos na Tabela 9. O colaborador 1 viajou por três zonas, portanto, gerou três registros, já o Colaborador 2, também viajou por 3 zonas e também gerou três registros. Este agrupamento de registros facilita a contagem da quantidade de viagens por zona (índice de mobilidade),

bem como possibilita estudar sua variação durante a semana, que é o objetivo desta dissertação.

Tabela 8 - Exemplo de agregação Pt. 1

Colaborador	Updated	Área
1	13:05:23	1
1	13:05:23	1
1	13:05:23	1
1	13:05:23	2
1	13:05:23	2
1	13:05:23	3
1	13:05:23	3
1	13:05:23	3
2	21:38:55	1
2	21:38:55	1
2	21:38:55	3
2	21:38:55	4
2	21:38:55	4

Foi realizado o agrupamento de registros com mesmo campo ‘updated’ e mesma Área, o resultado é 1 registro por área que o colaborador viajou, conforme Tabela 9.

Tabela 9 - Exemplo de agregação Pt. 2

Colaborador	Updated	Área
1	13:05:23	1
1	13:05:23	2
1	13:05:23	3
2	21:38:55	1
2	21:38:55	3
2	21:38:55	4

4.2.5. Contagem dos dados

Após o agrupamento, o conjunto de dados guarda informações acerca de cada área onde um dado colaborador viajou. Para o cálculo do índice de mobilidade faz-se necessário realizar a contagem dos registros por cada área, isto se deve ao fato de que cada registro representa uma viagem, conforme explanado no subitem anterior. Retomando o exemplo anterior, o resultado da contagem de ambos os colaboradores está disposto na Tabela 10.

Tabela 10 - Cálculo do Índice de Mobilidade

Área	Viagens ou Índice de Mobilidade
1	2
2	1
3	2
4	1

A Tabela 10 indica o índice de mobilidade, ou seja, o número de viagens efetuadas por zona. Considerou-se no índice de mobilidade a totalidade de áreas onde o colaborador esteve em determinado momento do dia, e não apenas sua origem e destino. Esta consideração torna as conclusões mais confiáveis, pois se o índice de mobilidade é um indicativo da demanda da rede de transporte, então a rede de transporte de pontos intermediários a uma viagem devem ser considerados, sob pena de subestimar a quantidade de viagens em determinadas zonas, as quais servem sobretudo a transbordos, por exemplo.

A variação do índice de mobilidade é a variação da intensidade da demanda à rede de transporte nos vários dias da semana. Portanto, se os colaboradores tem frequência diária rotineira durante a semana, então espera-se um índice estável, sem oscilações. Já se os colaboradores tem frequência instável, realiza viagens em alguns dias e em outros não, então o índice será afetado. O índice trata do impacto em dias distintos, considerando todas as possíveis compensações que podem ocorrer devido à comportamentos não rotineiros de vários colaboradores. Tratar do impacto na rede de transporte por meio de um índice possibilita maior simplicidade na aplicação de políticas de transporte.

4.3. Análise dos dados

Ao término do procedimento de tratamento de dados, espera-se ter uma tabela com o número de viagens por zona, como exemplificado pela Tabela 10. Conforme disposto no subitem referente a obtenção de dados, o download deverá ser realizado por no mínimo 15 dias, intervalo mínimo para observação da variação semanal, se houver. Para cada um desses dias deve ser realizado o procedimento de tratamento disposto neste capítulo. Com isso, tem-se que cada dia terá um índice de mobilidade e estudar se a variação deste índice é significativa ou não é o objetivo da análise de dados.

Para iniciar os estudos comparativos é necessário que se tenham dias representativos de cada dia da semana. Para isso, adotou-se a mediana dentre todos os índices de cada dia da semana. A mediana foi utilizada no intuito de desconsiderar os

pontos ‘fora da curva’ que foram observados nos índices de mobilidade. Os dias representativos estão dispostos no ANEXO II.

Para o estudo da variação do índice de mobilidade foi feito o uso dos teste estatísticos t-pareado e Analysis of Variance (ANOVA). A ANOVA foi utilizada para análise do índice de mobilidade entre os cinco dias úteis da semana (segunda a sexta). Após a análise de variância escolheu-se um dia útil representativo, recomenda-se que seja a média dentre os dias da semana (segunda a sexta), isto se a ANOVA não demonstrar grandes disparidades. O teste t-pareado foi utilizado na comparação entre os índices de mobilidade dos dias representativos e finais de semana ou feriados.

O método explicado e exemplificado neste capítulo está resumido no fluxograma da Figura 17.

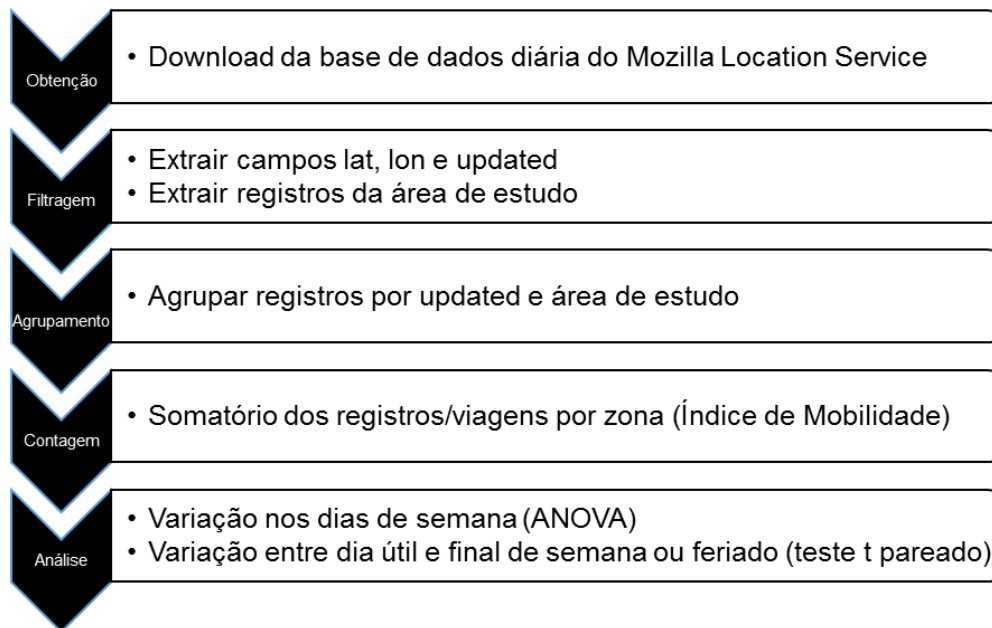


Figura 17 - Fluxograma do método

O capítulo seguinte aborda o estudo de caso realizado na cidade do Rio de Janeiro.

5. ESTUDO DE CASO

Neste capítulo foi realizado o estudo de caso para a cidade do Rio de Janeiro. O método proposto foi aplicado a um total de 86 bases de dados diárias do MLS, referentes aos meses de Maio, Junho e Julho. O procedimento de filtragem utilizou as coordenadas que cobrem a área da cidade do Rio de Janeiro, no intuito de obter todos os registros desta área de estudo. Além disso, apenas as colunas referentes ao horário de upload e coordenadas geográficas foram estudadas, os dados restantes foram excluídos.

O Rio de Janeiro é dividido por bairros, Áreas Administrativas e Áreas de Planejamento, sendo que as últimas são as maiores áreas dentre as divisões, totalizando-se em cinco. Fez-se uso das APs para o estudo dos dados do Mozilla Location Service, uma vez que, devido à grande área, os erros provenientes de imprecisões aos dados utilizados tendem a ser englobados ou ‘diluídos’.

A agregação e contagem foram realizadas por meio do software MySQL, tal software possibilitou grande eficiência no manuseio das 86 bases de dados, visto que todas as bases foram manuseadas em conjunto. Estes procedimentos resultam no índice de mobilidade.

5.1. Área de estudo

A cidade do Rio de Janeiro teve como origem o Porto na Baía de Guanabara. Com o passar do tempo, o que era a totalidade da cidade foi se transformando no Centro, local de negócios e administração, o uso de solo residencial era aos poucos substituído, se deslocando para os bairros e subúrbios. Num segundo momento as atividades econômicas se difundiram pela cidade, o que levou ao surgimento de diversas áreas comerciais fora do centro tradicional (GEISER, 2002).

Atualmente, o município tem mais de 6,3 milhões de habitantes, a segunda metrópole do país em termos de população e produto. A área total do município é de 1224,56 km² com diversas peculiaridades resultantes de sua geografia, tal como a maior floresta urbana do mundo, a Floresta da Tijuca. O município é dividido em quatro regiões geográficas comumente conhecidas como: Centro, Zona Norte, Zona Sul e Zona Oeste (PREFEITURA DO RIO DE JANEIRO, 2009, 2011?; GEISER, 2002).

As divisões por citadas acima são termos utilizados comumente pelos habitantes da cidade. Entretanto, a administração municipal, visando melhor coordenação e planejamento, realizou outra divisão setorial, foram criadas as Áreas de Planejamento (Figura 18).

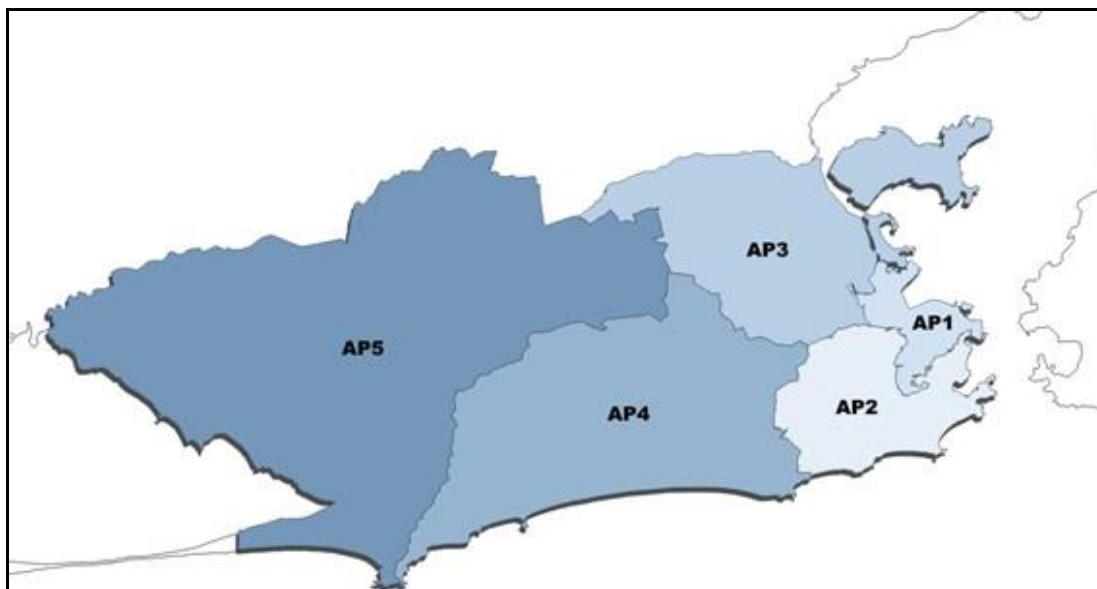


Figura 18 - Áreas de Planejamento do Rio de Janeiro

Fonte: Prefeitura do Rio de Janeiro (2015)

Cabe ressaltar que há menores divisões, tais como regiões administrativas, conhecidas por RA, além dos bairros, nível geográfico mais detalhado (PREFEITURA DO RIO DE JANEIRO, 2011?). Neste trabalho optou-se pela abordagem dos dados a nível de AP, a fim de reduzir os erros que as imprecisões dos dados poderiam acarretar. Conforme pode ser visualizado na Figura 18, existem 5 APs na cidade do Rio de Janeiro. Tais APs, também denominadas macrozonas, são detalhadas a seguir.

5.1.1. Área de Planejamento 1

Formada por 15 bairros, com 4,7% da população carioca – 297.976 habitantes, segundo Censo 2010. Com território de 34,39% km², esta AP abrange cerca de 2,8% do território do município e tem densidade populacional de 8.665 habitantes por km². No período de 2000 a 2010, a população desta macrozona cresceu em 11% (PREFEITURA DO RIO DE JANEIRO, 2007?, 2011).

É o centro histórico da cidade, mas também a área que mais sofreu transformações do cenário urbano, as atividades econômicas foram se superpondo em detrimento das funções residenciais. Entre os anos de 1970 e 2000 chegou a perder quase 27% de sua população, mas atrai diariamente quase um milhão de pessoas que trabalham nessa região, principalmente no setor de serviços (PREFEITURA DO RIO DE JANEIRO, 2007?; Geiser, 2002).

Esta área é também uma zona de passagem para circulação intra ou interurbana. Nela se localizam o terminal rodoviário Novo Rio, o terminal ferroviário da

Central com integração ao Metrô e diversas linhas de ônibus na Presidente Vargas, O aeroporto doméstico Santos Dumont, o Veículo Leve sobre Trilhos (VLT) carioca, a estação das barcas para Niterói, Ilha do Governador e Ilha de Paquetá, entre outras infraestruturas voltadas ao transporte (GEISER, 2002).

5.1.2. Área de Planejamento 2

Contém 25 bairros, representando 16% dos cariocas – 1.009.170 habitantes, segundo Censo 2010. Ocupa uma área de 100,43 km² que corresponde a 17% do município, a densidade populacional é de 10.048 habitantes por km². O aumento populacional entre o Censo de 2000 e 2010 foi de apenas 1,2%, número que mostra o quanto a região é consolidada, além do fato de a geografia minimizar expansões (PREFEITURA DO RIO DE JANEIRO, 2007?, 2011).

Corresponde a uma periferia próxima ao Centro, na direção sul e sudoeste. Com configuração geográfica entre o mar e o maciço da Tijuca, é a região que simboliza a imagem da cidade para o mundo, é chamada Zona Sul da cidade. O termo zona sul ganhou significado valorativo, haja vista que abriga a maior parte das classes altas da cidade, apresentando altos padrões de infraestrutura e de serviços, além de um quadro de amenidades naturais (PREFEITURA DO RIO DE JANEIRO, 2007?; GEISER, 2002). Na AP2 está localizado a Linha 1 do metrô, que faz a ligação entre os bairros da Tijuca e Ipanema, bem como a Linha 4, que liga Ipanema à Barra da Tijuca.

5.1.3. Área de Planejamento 3

Possui 80 bairros que correspondem a 16,6% do território municipal – 203,47 km². Engloba 40,2% do total da população residente no Rio de Janeiro – 2.398.572 habitantes, segundo Censo de 2010, o que resulta numa densidade populacional de 11.788 habitantes por km². É a AP mais populosa e assim como a AP2 é bastante consolidada, uma vez que de 2000 a 2010 o aumento populacional chegou a cerca de 2% (PREFEITURA DO RIO DE JANEIRO, 2007?, 2011).

A AP3 é formada pelas planícies que se estendem ao norte do maciço da Tijuca até os limites da cidade na baixada fluminense. É um mosaico em termos de composição social em que prevalecem camadas populares e de baixa classe média, com destaque para grandes áreas de favelização, tais como a Favela da Maré e o Complexo de Favelas do Alemão (GEISER, 2002). Esta macrozona também agrega duas ilhas: A

Ilha do Governador, onde está localizado o Aeroporto Internacional Antônio Carlos Jobim e a Ilha do Fundão onde fica a Universidade Federal do Rio de Janeiro (UFRJ).

Quanto ao transporte, o eixo central é rodoviário formado pelas rodovias Avenida Brasil e Linha Vermelha, as principais vias das relações interestaduais da cidade. A Linha Amarela é outra importante via expressa da cidade que liga a Barra da Tijuca à Ilha do Fundão.

5.1.4. Área de Planejamento 4

É formado por 19 bairros. No conjunto da cidade, a AP4 corresponde a 24% da área – 293,79 km² e 14,4% da população carioca – 909.955 habitantes, segundo o Censo 2010. Sua densidade populacional é de 3.097 habitantes por km². Apresentou o maior crescimento populacional dentre todas as APs, 33%, número que evidencia o quanto esta região está em franca expansão (PREFEITURA DO RIO DE JANEIRO, 2007?, 2011).

É uma imensa praia oceânica, mais de 25 km, se estende na metade ocidental do município, é limitada pelos maciços da Tijuca e Pedra Branca e pelo oceano atlântico. Tal região viu nos anos 70 o início de um processo acelerado de urbanização, compreendendo setores de população de alta classe (GEISER, 2002). A Prefeitura do Rio de Janeiro [2007?] destaca que nesta região existe grande dependência do transporte individual, entretanto, esforços estão sendo realizados para induzir o uso do transporte público. Nesse sentido destaca-se o *Bus Rapid Transit* (BRT) Transoeste que liga o bairro Jardim Oceânico à Santa Cruz, com 56 km de extensão, e também a Linha 4 do metrô que liga a zona sul à barra.

5.1.5. Área de Planejamento 5

Formada por 20 bairros tem a maior extensão territorial dentre as APs – 592,45 km², totalizando 48,4% do território do município. É a segunda área mais populosa dentre as Áreas de Planejamento. Abriga cerca de 27% da população carioca – 1.704.773 habitantes, segundo Censo 2010, sendo que comparado ao ano 2000 houve um crescimento de 9,5% da população desta AP. Sua densidade populacional é de 2.877 habitantes por km² (PREFEITURA DO RIO DE JANEIRO, 2007?).

É uma área na qual a cidade ainda pode expandir, haja vista a presença de vazios urbanos suscetíveis à urbanização, os recursos ambientais, a implantação de grandes empreendimentos e a interação com municípios limítrofes. É chamada,

habitualmente, de zona oeste, apesar de, na prática, a zona oeste também incluir a Barra da Tijuca (AP4). A denominação, por meio do uso popular, passou a identificar somente a AP5 e alguns bairros da AP4, com características socioeconômicas parecidas com a da AP5 (PREFEITURA DO RIO DE JANEIRO, 2007?, 2011?).

Esta imprecisão na nomenclatura é resultado da estreita relação entre a localização espacial e nível social, uma peculiaridade da cidade, onde características geográficas são generalizadas e confundidas com caracterização social. A AP5 é a que tem maior participação relativa nos transportes coletivos, para transporte individual a participação é de 12,9% e 87,1% para o coletivo, já a porcentagem do município são, respectivamente, 28% e 72%. Além disso, os habitantes desta AP ainda precisam atravessar a região mais densa em trânsito, a AP3 e parte da AP1 para chegar ao Centro da cidade, local de trabalho de muitos residentes da área (Prefeitura do Rio de Janeiro, 2011?).

Na Tabela 13 é feito um resumo para caracterizar cada uma das APs do Rio de Janeiro. No ANEXO I estão dispostos os bairros por AP.

5.2. Obtenção dos dados

O Mozilla Location Service disponibiliza bases de dados diariamente. Tais bases ficam disponíveis no site do MLS por, no máximo, três dias, após este período o download fica impossibilitado. Os arquivos contêm basicamente um aglomerado textual no qual cada palavra é separada por vírgula, o que caracteriza o formato .csv (Comma Separated Values). Tem aproximadamente 500Mb e 22 milhões de registros por base de dados.

Estes arquivos armazenam localizações de antenas telefônicas de várias partes do mundo. Como já explicado anteriormente, estas localizações são atualizadas pelos colaboradores por meio da coleta e upload de dados do app Mozilla Stumbler. As localizações atualizadas, bem como o horário de upload são suficientes para extrair o deslocamento aproximado do colaborador.

Portanto, a base de dados é composta por localizações de antenas, localizações que foram atualizadas tanto no dia de disponibilização da base, como também anteriormente. Este trabalho fez uso dos dados de localizações referentes ao dia em que a base de dados foi fornecida, ou seja, serão descartados os dados atualizados em datas anteriores ao dia de disponibilização.

Tabela 11 – Descrição das APs do Rio de Janeiro

AP1	AP2	AP3	AP4	AP5
<p>Principal centralidade da cidade, o Centro é o bairro com maior influência na região metropolitana</p> <p>Área de ocupação antiga, abrangendo o Centro Histórico da Cidade com a presença de várias áreas protegidas (APACs) e bens tombados</p> <p>Malha viária de alcance metropolitano, tais como: Av. Brasil, Linha Vermelha e Av. Presidente Vargas.</p>	<p>Principal referência da imagem da Cidade em nível nacional e internacional</p> <p>Área próxima ao Centro da Cidade e a Barra da Tijuca. Abriga pontos turísticos da Cidade</p> <p>Malha estruturadora dos eixos Norte-Sul-Centro da cidade e Barra da Tijuca</p>	<p>Proximidade com a região metropolitana</p> <p>Principal porta de entrada da cidade – Av. Brasil e Aeroporto Internacional Antônio Carlos Jobim</p> <p>Malha viária de alcance metropolitano destacando-se a Av. Brasil, as linhas Amarela e Vermelha</p>	<p>Área de expansão do município com grande aumento de moradores</p> <p>Grande baixada, circundada pelos maciços da Tijuca e Pedra Branca e com grande parte da sua área ainda preservada</p> <p>Ao longo da Av. das Américas estão dispostos os shoppings, fazendo desta área um Centro de Comércio de alcance metropolitano</p>	<p>Proximidade com a região metropolitana</p> <p>Eixos estruturadores com o município e a região metropolitana – Av. Brasil, BR 101, Av. das Américas e Estrada Rio/São Paulo</p> <p>Principais corredores de transporte de massa – Av. Brasil e rede ferroviária, ramal de cargas e passageiros</p>

Fonte: Prefeitura do Rio de Janeiro [2007?]

Fez-se uso de 86 bases de dados, entre os meses de Maio, Junho e Julho, incluindo os feriados do dia do trabalho (1º de maio) e Corpus Christi (15 de junho). Alguns dias foram perdidos, entretanto, para cada dia da semana (segunda a domingo) foram coletados ao menos 11 dias, conforme pode ser observado no ANEXO II. Além disso, os 86 dias superam com folga a recomendação de 15 dias, fornecida por Huff e Hanson (1988), no que se refere a quantidade de dias necessários ao estudo da variação de padrões de viagem.

5.3. Filtragem dos dados para o Rio de Janeiro

Em cada uma das 86 bases de dados, apenas as colunas lat e lon, referentes às coordenadas geográficas, e a coluna updated foram utilizadas. As coordenadas foram utilizadas a fim de obter o deslocamento aproximado, já o horário de upload foi usado para identificar um colaborador.

Após o descarte das colunas não utilizadas restarão três colunas, entretanto, 22 milhões de linhas em cada arquivo. Filtra-se tais registros com base nas coordenadas geográficas do Rio de Janeiro, tais coordenadas formam uma área retangular que engloba praticamente toda a área do município do Rio de Janeiro, conforme Figura 19.

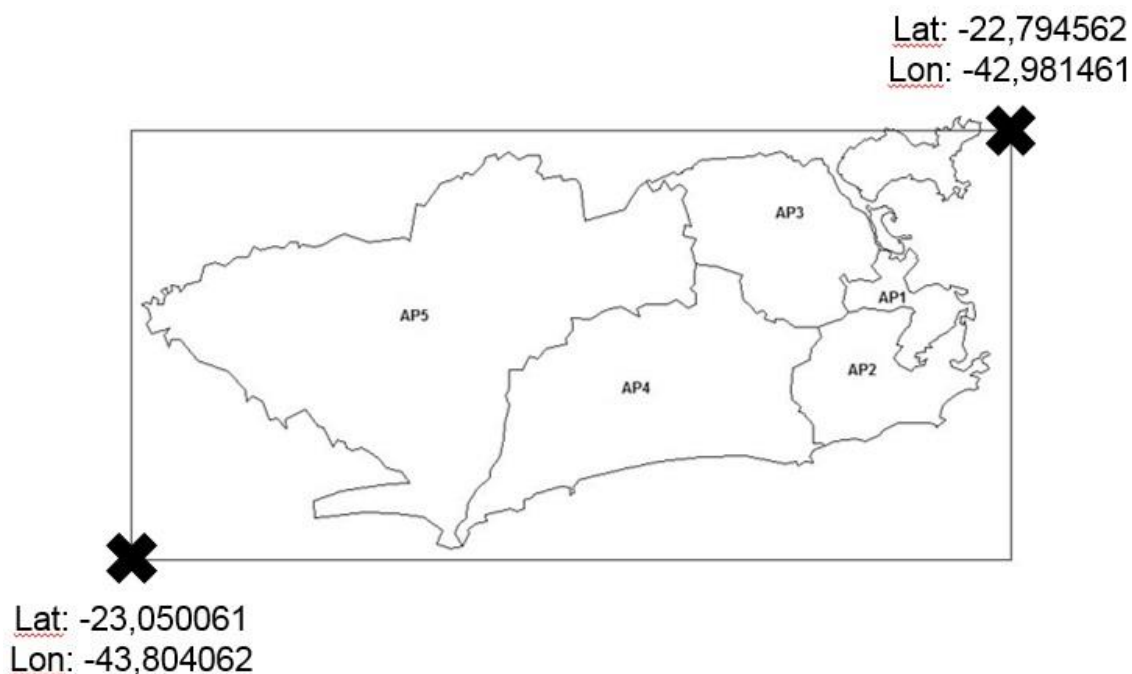


Figura 19 - Coordenadas do Rio de Janeiro

Nota-se, na Figura 19, que a área retangular abrange regiões além daquelas referentes à cidade de estudo, tal como partes da baixada fluminense. Tais regiões serão descartadas no passo seguinte, referente às atribuições das áreas a cada ponto. Ao final da

filtragem os registros somam um total de 73 mil em média por dia, eles se referem às atualizações das localizações de diversas antenas na cidade do Rio. Infelizmente, o número diário de 73 mil não se trata de todos os registros gerados em um dia de detecção, mas sim a junção de todas as últimas observações (última atualização) de antenas que podem ter sido atualizadas no dia em que foi feito o *download* ou anteriormente a isso, conforme já mencionado no subcapítulo anterior.

Os registros de atualização diária para a cidade do Rio de Janeiro, e que serão utilizados por este trabalho, ficam entorno de 2000.

5.4. Atribuição de áreas para cada registro

As áreas consideradas para atribuição dos registros são as Áreas de Planejamento estabelecidas pela Prefeitura do Município. Tais regiões tem formas que podem ser aproximadas por um determinado número de retângulos, variando de acordo com a complexidade da forma e tamanho da AP, conforme Figura 20. A aproximação foi realizada com precisão acima da adotada no processo de filtragem (Figura 19), entretanto, devido ao fato de que a representação exata das APs demandariam vários retângulos, optou-se por utilizar a aproximação da Figura 20.

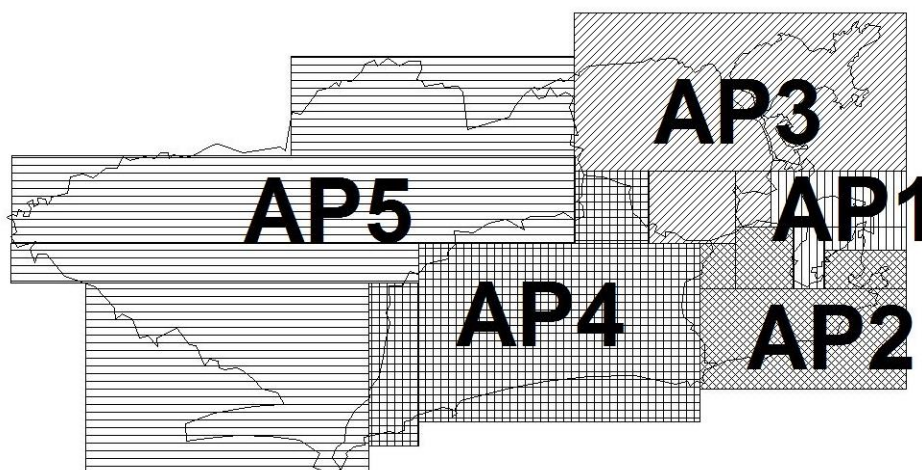


Figura 20 - Aproximação das APs em retângulos

Tal precisão possibilitou o descarte de alguns registros referentes à áreas da baixada fluminense. De posse das coordenadas geográficas de cada retângulo representado na Figura 20, a atribuição de áreas foi realizada por meio de fórmulas “Se... Então...” pelo software Microsoft Excel. Funciona da seguinte maneira, elaboram-se as fórmulas considerando as coordenadas de cada retângulo das respectivas Áreas de

Planejamento. Tais fórmulas tem a tarefa de verificar se a longitude, na coluna lon, está dentro do intervalo horizontal limitado pelos retângulos, bem como verificar se a latitude, na coluna lat, está dentro do intervalo vertical limitado pelos retângulos. Se ambas as condições foram obedecidas, então a fórmula retornará o valor VERDADEIRO, se ao menos uma das condições for falsa, o valor retornado é FALSO (Tabela 12).

Tabela 12 - Atribuição de valores Pt. 1

lon	lat	updated	AP5	AP4	AP3	AP2	AP1
-43,106	-22,8833	10:18:09	FALSO	FALSO	FALSO	FALSO	VERDADEIRO
-43,0899	-22,8425	19:31:24	FALSO	VERDADEIRO	FALSO	FALSO	FALSO
-43,091	-22,8437	05:26:46	FALSO	FALSO	VERDADEIRO	FALSO	FALSO
-43,0262	-22,8069	05:26:46	FALSO	FALSO	VERDADEIRO	FALSO	FALSO

Após o procedimento do parágrafo acima resta adicionar uma coluna à tabela contendo as áreas onde os pontos estão contidos. Para isto a mesma fórmula “Se... Então...” foi utilizada. Por exemplo, se o registro da coluna AP5 apresentar valor VERDADEIRO, então a coluna Local apresentará o valor 5, caso contrário, verifica-se a coluna AP4 e assim por diante. Caso não haja valor VERDADEIRO em nenhuma das APs significa que o ponto não pertence aos retângulos formulados, muitas vezes tais pontos foram referentes à baixada fluminense.

A Tabela 13 demonstra o quadro resultante da solução adotada.

Tabela 13 - Atribuição de APs para cada colaborador

lon	lat	updated	AP5	AP4	AP3	AP2	AP1	Local
-43,106	-22,8833	10:18:09	FALSO	FALSO	FALSO	FALSO	VERDADEIRO	1
-43,0899	-22,8425	19:31:24	FALSO	VERDADEIRO	FALSO	FALSO	FALSO	4
-43,091	-22,8437	05:26:46	FALSO	FALSO	VERDADEIRO	FALSO	FALSO	3
-43,0262	-22,8069	05:26:46	FALSO	FALSO	VERDADEIRO	FALSO	FALSO	3

5.5. Agregação dos dados

Um colaborador do MLS gera vários registros em uma mesma área quando é feito o uso do app Mozilla Stumbler. No capítulo acerca do método proposto mostrou-se que apenas 1 registro do colaborador por área visitada é suficiente para inferir que tal colaborador realizou uma viagem na área considerada. Tal conclusão se deve ao fato de que o app gera registros apenas sob movimento.

O agrupamento de registros reduz o tamanho do banco de dados e facilita a contagem de viagens por zona, que é o índice de mobilidade. O código utilizado é mostrado no Anexo III.

5.6. Contagem das viagens efetuadas

Devido às características dos dados, o comportamento semanal para um único indivíduo não pode ser observado, com isso, a estratégia adotada por este estudo é utilizar um índice que resulte na síntese das viagens realizadas por todos os colaboradores em um dia qualquer. Adjetiva-se tal valor de ‘indicador agregado’, pois não trabalha a nível de indivíduo, mas sim considera um comportamento resultante diário dos deslocamentos de todos os colaboradores.

Ao número que condensa os comportamentos de todos os colaboradores denomina-se índice de mobilidade. O índice de mobilidade é o número que traduz a intensidade da mobilidade dos colaboradores intra e entre APs. Há a possibilidade de inferir quais são os dias mais ou menos movimentados de acordo com o índice, o que possibilita um parâmetro prático para a tomada de decisões.

O índice foi calculado por meio da contagem de registros por zona. Uma vez que os registros foram agrupados, na base de dados restaram apenas 1 registro de um dado colaborador pelo zona onde o mesmo realizou sua viagem. Portanto, o somatório dos registros resulta na quantidade de viagens realizadas na zona de estudo, o índice de mobilidade.

5.7. Análise estatística dos índice de mobilidade

Após o cálculo do índice de mobilidade para cada uma das bases obtidas, o último passo é a análise estatística dos números obtidos. Antes, algumas medidas foram tomadas no intuito de adequar os dados aos testes t-pareado e ANOVA, que foram os testes estatísticos utilizados neste trabalho. A primeira medida foi obter um dia representativo para cada dia da semana, ou seja, índices de mobilidade para segunda a domingo. O dia representativo foi obtido pela mediana dentre os vários dias coletados, escolheu-se a mediana pela característica de minimizar os efeitos dos *outliers*, isto é, os valores atípicos.

Os dias úteis representativos foram analisados por meio da análise de variância. Após o estudo entre os dias úteis foi realizada a análise considerando o comparativo entre dia útil-fim de semana e dia útil-feriado. Em ambos os casos foi utilizado a média

entre os dias úteis representativos na análise entre sábado, domingo e os feriados de Corpus Christi e Dia do Trabalho.

Os resultados dessa análise estão dispostos no capítulo 6.

6. RESULTADOS E DISCUSSÕES

Neste capítulo foi feita a apresentação e análise dos resultados obtidos pelo estudo da variabilidade com dados do MLS. No primeiro subcapítulo a análise é realizada graficamente, no intuito de gerar uma visão geral da mobilidade nas APs, bem como propiciar conhecimentos prévios relativos à variação que ocorre tanto entre os dias da semana, quanto entre as macrozonas.

No subitem 4.3 apresenta-se a variabilidade entre os dias da semana, procura-se responder se existem variações estatisticamente significativas no índice de mobilidade entre os dias da semana. Em 4.4 é realizada o estudo comparativo entre a mobilidade do dia útil e feriado, buscando constatar em qual nível a mobilidade do dia útil difere de sábado e domingo. Em 4.5 buscou-se realizar a comparação de dia útil e feriados, dois feriados foram analisados, dia do trabalho e Corpus Christi, foi feita também uma comparação entre estes dois feriados. Por fim, apresenta-se uma recomendação de uso dos dados do MLS em conjunto com o app Mozilla Location Service, no intuito de aprimorar os resultados obtidos.

6.1. Análise gráfica preliminar

A primeira análise deste trabalho reside no estudo do índice de mobilidade por AP. A maneira mais apropriada para demonstração de tais resultados, seguindo Sagl et al. (2012), é um gráfico radar conforme os gráficos de 1 a 5. Para AP1 dois aspectos são notáveis, o primeiro é que na sexta-feira há maior IM entre os dias úteis, resultado que está em consonância com Dharmowijoyo (2015).

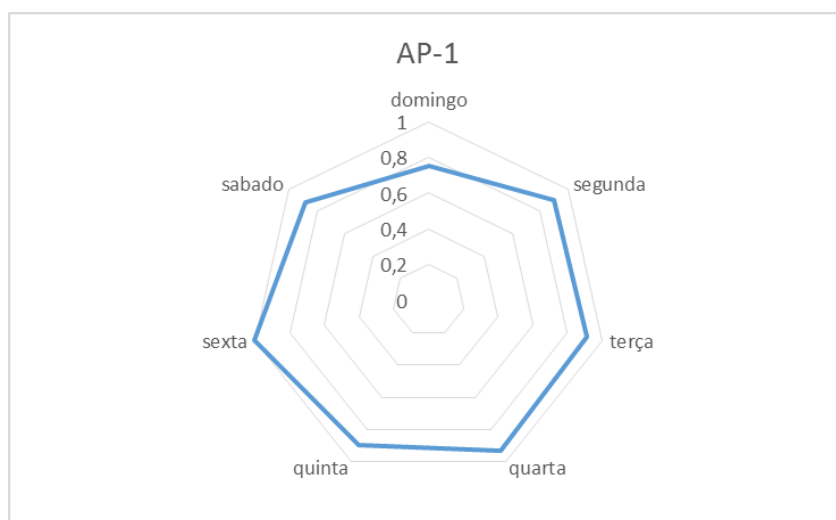


Gráfico 1 - Índice de mobilidade relativo para AP1

A AP1 é composta por centros comerciais, entretanto, há também uma grande quantidade de ambientes de socialização, tais como, bares e boates. Nesta macrozona se localiza o bairro da Lapa, mundialmente conhecido pelo entretenimento noturno. Estes fatores tendem a aumentar o número de viagens e o alto índice de mobilidade é um indicativo de que a amostra segue tal tendência. O segundo aspecto é o menor IM no domingo, resultado esperado devido ao fato de domingo ter mobilidade bem menos intensa em centros comerciais.

Ainda em AP1, percebe-se uma redução no IM no sábado, mas não tão acentuada quanto domingo, é possível que isso seja resultado de que parte da amostra tem uma ocupação neste dia da semana, seja profissional ou relativo a compras. A redução na segunda-feira é recorrente em todas as APs, sendo mais perceptível nas AP1 e AP2. Ahas et al. (2010) foram os únicos, dentro da literatura abordada, a citar o possível impacto que o final de semana gera nos dias úteis próximos a eles, ou seja, segunda e sexta. Os autores não citaram quais os impactos possíveis, apenas citaram a ideia, os dados obtidos pelo MLS resultam que a influência do final de semana na sexta-feira é no sentido de aumentar o IM, do contrário da segunda-feira, na qual a tendência é diminuir.

No gráfico 2, percebe-se que assim como na AP1, o IM é menor aos domingos. Este resultado é discordante ao que se imagina ser recorrente na zona sul, uma vez que é a área com diversos pontos turísticos e amenidades, que geralmente são visitadas aos domingos. Entretanto, para a amostra estudada esta tendência não foi observada.

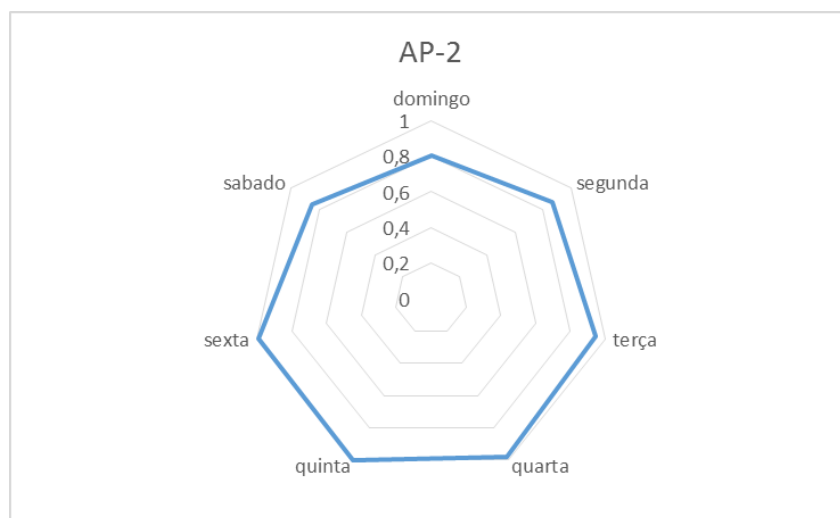


Gráfico 2 - Índice de mobilidade relativo para AP2

Novamente os índices de sexta e segunda diferem dos demais dias úteis, sendo que sexta tem o maior índice no geral e segunda, o menor entre dias úteis. A hipótese de

Ahas et al. (2010) de que os finais de semana tem influência direta nos dias úteis próximos a eles, também é observada na AP2. Supõe-se que tamanha similaridade com AP1, se deve ao fato de que AP2 também é um grande gerador de empregos, complementar ao Centro, e devido a isso, a frequência de viagens tende a ter menor variabilidade (PAS e KOPPELMAN, 1986) que em áreas com características diferentes.

A AP3 difere das anteriores pelo fato de sábado ser o dia de maior IM (Gráfico 3). Possivelmente o costume de ir às compras ao sábado, também destacado por Ahas et al. (2010), pode ter interferido nos dados obtidos para esta macrozona. Além disso, o fato de sábado e sexta-feira terem comportamentos iguais e distintos de outros dias úteis está em consonância aos achados de Susilo e Kitamura (2005), que, assim como nesta AP, encontrou que domingo tem menor IM em comparação com todos os outros dias.

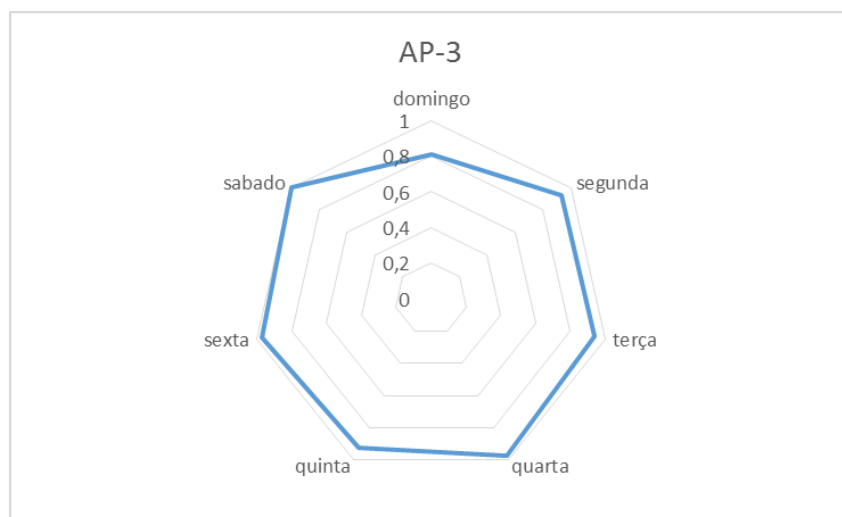


Gráfico 3 - Índice de mobilidade relativo para AP3

Na AP4 há grande redução de mobilidade na quinta e sexta, sendo que nestes dias o índice é apenas um pouco maior que domingo, conforme Gráfico 4. Observa-se que o dia com maior intensidade de mobilidade é terça-feira. Um comportamento tão heterogêneo entre os dias de semana não era esperado, neste ponto pressupõe-se que a quantidade de dados disponíveis na região afeta substancialmente as informações resultantes. No Anexo II pode-se observar a diferença que existe entre o número de registros, e conseqüentemente de colaboradores, existente entre a AP4 e, por exemplo, a AP2 e AP3, que apresentaram maiores graus de similaridade entre os dias de semana. A AP2 tem 50% mais registros, já a AP3 tem mais que o dobro, em comparação com a macrozona que representa a região da Barra da Tijuca.

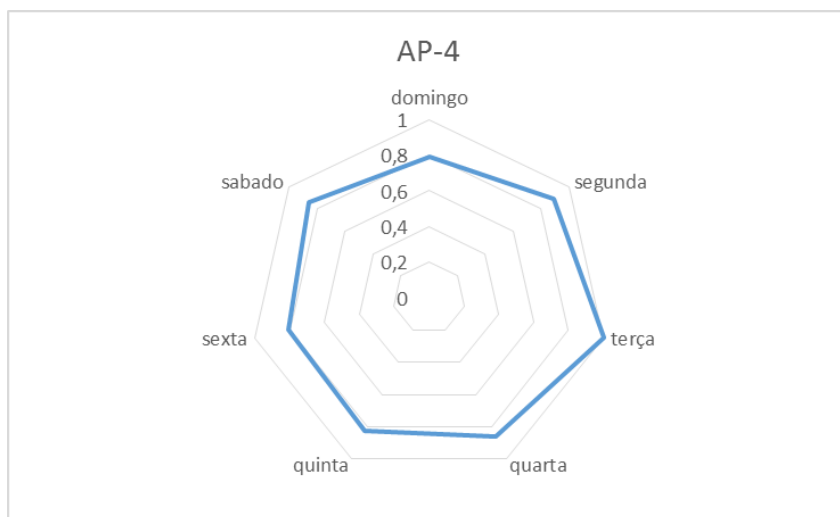


Gráfico 4 - Índice de mobilidade relativo para AP4

Observa-se que em AP5, ainda que o gráfico tenha apresentado maior disparidade diante dos demais, a principal característica foi mantida, isto é, o achatamento do gráfico no domingo. Além disso, assim como em AP3, o maior IM é de sábado. Reitera-se a conjectura de que, para os colaboradores desta Área de Planejamento, o sábado pode ser um dia no qual é habitual a realização de compras de produtos que podem ser obtidos por meio de viagens dentro da zona. A dissimilaridade entre os dias de semana também é notável, e igualmente a AP4, o baixo número de colaboradores pode ser a razão da distorção observada.

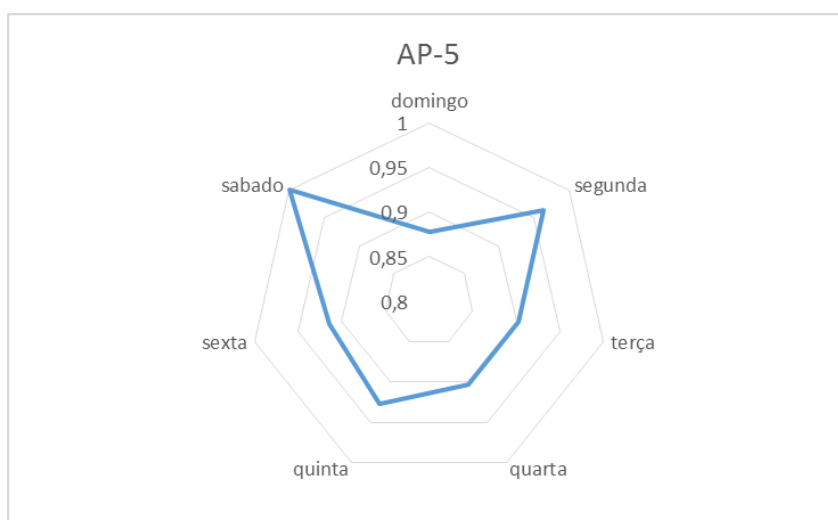


Gráfico 5 - Índice de mobilidade relativo para AP5

A baixa intensidade de mobilidade encontrada com dados do Mozilla Location Service para o dia de domingo está de acordo com diversos autores na literatura pesquisada, por exemplo, conclusões são encontradas neste sentido em Dharmowijoyo (2015) e Raux et al. (2015). Alguns autores ainda destacam que o domingo é voltado às

atividades sociais (ALLSTROM et al., 2016), o que pode ou não demandar viagens, o que explica a redução do IM aos domingos, mas não em um grau tão alto.

Quanto ao dia de sábado os resultados se diferiram entre as macrozonas estudadas. Nas APs 1, 2 e 4 houve redução de mobilidade da amostra, enquanto que nas APs 3 e 5 o IM foi máximo. As macrozonas em que ocorreram reduções estão de acordo com os achados de Raux et al. (2015), estes autores encontraram que tal redução acontece em menor grau que no domingo, o que ocorre também nas APs estudadas. Já nas APs de mobilidade máxima há proximidade relativa com as conclusões obtidas por Susilo e Kitamura (2005) e Allstrom et al. (2016), sendo que o último destaca haver uma tendência de que viagens a compras são feitas majoritariamente aos sábados, o que afeta o IM calculado por este estudo.

Os dias de semana apresentam basicamente índices de mobilidade mais constantes, sempre maiores que domingo. Na segunda observa-se uma redução nas APs 1, 2 e 3, já nas APs 4 e 5 há maior disparidade entre os dias da semana, talvez tal disparidade possa vir da menor quantidade de dados provenientes destas APs em comparação com as três anteriormente citadas.

Esta primeira análise baseou-se na observação de gráficos dos índices de mobilidade relativos. Tal método é importante no sentido de introduzir os resultados, entretanto, não fornece a credibilidade de uma análise numérica. Com isso, nos próximos itens deste capítulo será dada ênfase à análise de dados baseado em estatística, a fim de proporcionar maior confiabilidade aos resultados obtidos.

6.2. Variabilidade entre dias úteis

A primeira análise estatística realizada por este estudo é a comparação entre os dias úteis representativos. Conforme mencionado em 3.5, os índices de mobilidade dos dias representativos foram obtidos pela mediana dos índices de mobilidade de uma total de 12 dias para cada dia da semana, isto foi feito no intuito de minimizar os efeitos dos *outliers* na análise dos dados.

Desconsideraram-se, no cálculo dos dias representativos, o feriado de 1º de maio, dia do trabalho que caiu numa segunda-feira, e também o feriado de Corpus Christi, que no ano de 2017 aconteceu no dia 15 de junho, quinta-feira. Tais dias tiveram uma análise em particular, a consideração dos mesmos na obtenção dos dias representativos poderia desviar os valores obtidos. Tais dias estão dispostos na Tabela 8.

Tabela 14 - Dias representativos

	Segunda	Terça	Quarta	Quinta	Sexta
AP1	69	70	72	69	77
AP2	80	87	90	92	91
AP3	129	129	134	128	134
AP4	51	57	49	47	46
AP5	79	74	74	76	75

Neste subcapítulo pretende-se responder a seguinte pergunta: o dia útil influencia o índice de mobilidade nas APs? Para responder tal questão realizou-se uma análise de variância (ANOVA) com medidas repetidas, visto que as medidas não são independentes de um grupo a outro. A hipótese nula (H_0) deste teste estabelece que a média para os cinco grupos é a mesma, já a hipótese alternativa (H_1) estabelece que há ao menos uma diferença entre as médias dos grupos.

Neste item utilizou-se o valor-P como parâmetro de rejeição ou não da hipótese nula, nos itens seguintes fez-se uso também da comparação entre a estatística do teste e um valor crítico. O valor-P é frequentemente mais utilizado pois fornece, ao tomador de decisão, informações a respeito de o que o valor calculado da estatística do teste estava apenas nas proximidades da região de rejeição ou se estava muito longe dela (MONTGOMERY e RUNGER, 2009).

O teste realizado no software SPSS resultou em um valor-P de 0,761, o que leva a não rejeição da hipótese nula para um nível de significância de 5%. Portanto, rejeita-se a hipótese H_1 de que há diferenças significativas entre os grupos. Conclui-se, então, que o dia da semana não interfere no IM e que a resultante do comportamento dos colaboradores do MLS se mantém constante durante a semana, isto é, o impacto na rede de transporte tem baixa variabilidade.

Diante desse resultado, pode-se afirmar que, para esta amostra, a coleta de dados de um dia da semana não traria diferenças significativas nas previsões realizadas. Isso, possivelmente, se deve ao fato de a frequência das viagens, bem como as trajetórias são rotineiras, o que leva a uma baixa variabilidade nos dias de semana. Não se pode afirmar com alto grau de certeza a assertiva anterior, pois pode haver compensação entre comportamentos individuais, isto é, o indivíduo que não executou uma determinada trajetória no dia, pode ser compensado por outro indivíduo que também alterou seu padrão de viagem no dia selecionado. Tem-se aqui uma vantagem do estudo agregado perante os estudos realizados na literatura, pois considerar o

comportamento de todo um grupo irá mostrar a resultante da variabilidade na rede de transporte, que pode ter impacto nulo, embora a variabilidade individual seja alta.

Com isso, se a amostra estudada fosse representativa da população do Rio, o PDTU-Rio teria eficácia em sua modelagem de transporte, uma vez que para este Plano são coletados apenas as viagens realizadas no dia anterior à entrevista.

Cabe salientar aqui dois trabalhos, de Jones e Clarke (1988) e Schlich e Axhausen (2003), tais autores afirmam que a variabilidade aumenta quando o método tem a capacidade de capturar a complexidade do padrão de viagens. Os dados do MLS não fornecem um alto nível de profundidade de detalhamento dos dados, possibilitando apenas a obtenção do número total diário de viagens realizadas pelos colaboradores nas macrozonas do Rio. Também, devido ao fato de o horário de carregamento não ser, efetivamente, o horário de deslocamento, deve-se trabalhar com os dados do dia, desconsiderando as horas de upload. Portanto, não se considera estritamente o horário no qual o indivíduo realizou a viagem, o modo de transporte, o tempo de viagem, entre outros elementos que caracterizam quaisquer viagens. Devido a isso, a baixa variabilidade, ao menos em dias úteis, era esperada.

A estabilidade no comportamento de viagens encontra-se de acordo com Schlich e Axhausen (2003), estes autores estudaram o dia de sexta-feira no intuito de verificar a existência de variabilidade, encontrou-se uma alta correlação entre sexta e os outros dias úteis, um indicativo de baixa variabilidade, assim como o indício encontrado nesta dissertação.

Xianyu et al. (2015) calculou uma média de 2,6 viagens em dias úteis em sua amostra, como não foi realizada nenhuma observação acerca da variabilidade entre os dias úteis, deduz-se que tal variabilidade não é significativa e, portanto, constante de segunda a sexta-feira. Também existem discordâncias quanto ao achado deste subcapítulo, entre eles podem-se citar Dharmowijoyo et al. (2015, 2015b) que destacam o número mais alto de viagens na sexta-feira em comparação com outros dias úteis. Kang e Scott (2010) não observaram, em dias da semana, padrões de viagens uniformes de um dia para o outro.

As próximas análises são feitas com base na comparação de 1 dia útil representativo com os dias de final de semana, sábado e domingo, e dois feriados, dia do trabalho e Corpus Christi. Uma vez que a análise de variância concluiu que não há grandes diferenças entre os dias úteis, poder-se-ia utilizar um dia qualquer para representar um dia útil. Entretanto, isso poderia ocasionar em hesitação na escolha, por

este motivo optou-se por realizar os cálculos considerando a médias dos índices de mobilidade obtidos em dias úteis. Portanto, o dia útil utilizado nos cálculos posteriores é o grupo disposto na Tabela 9.

Tabela 15 - Dia útil representativo

Zona	Dia útil
AP1	71
AP2	88
AP3	131
AP4	50
AP5	76

Nos próximos subcapítulos foram realizadas análises comparativas.

6.3. Variabilidade entre dia útil e final de semana

De posse do dia útil representativo, neste subcapítulo foi realizada a análise estatística dos dados, comparando o dia útil com os dias de final de semana, isto é, sábado e domingo. Como é uma análise entre dois grupos de valores não independentes, o teste escolhido foi o t-pareado.

No teste t-pareado a hipótese nula estabelece que não há diferença entre as médias dos grupos analisados, já a hipótese alternativa tanto pode ser unilateral como bilateral, dependendo da conclusão a ser retirada se H_0 é rejeitada. Se o objetivo é fazer uma alegação envolvendo afirmações, tais como, maior que, menor que, superior a e assim por diante, uma alternativa unilateral é apropriada. Se nenhuma direção é implicada pela alegação, ou se for feita a alegação não igual a, uma alternativa bilateral deve ser usada (MONTGOMERY e RUNGER, 2009).

6.3.1. Comparação do IM entre dia útil e Sábado

O gráfico 6 mostra a comparação entre os índices de mobilidade de sábado e dia útil. É possível constatar a baixa diferença entre os dias analisados, além disso, em algumas APs o IM é maior aos sábados, em outras é maior no dia útil. Esta similaridade é evidenciada numericamente na Tabela 10, percebe-se que existem variações positivas e negativas no índice de mobilidade, no entanto, o somatório dessas variações é nulo.

O valor resultante nulo pode causar engano ao se tentar inferir os comportamentos das APs, pois pode-se imaginar que o IM não se alterou, entretanto, na

verdade existem variações consideráveis, sobretudo em AP2, mas que somadas aos índices de outras macrozonas resulta num valor nulo.

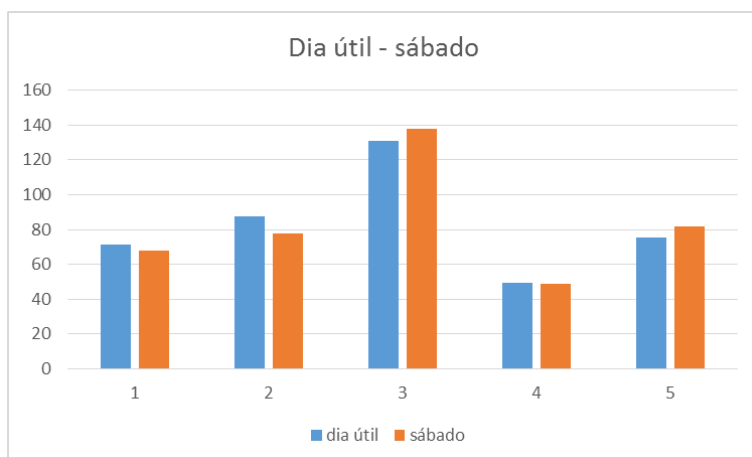


Gráfico 6 - Diferença entre índices de mobilidade - sab

Tabela 16 - Variação de IM entre sábado e dia útil

zona	dia útil	sábado	Δ
AP1	71	68	-4%
AP2	88	78	-11%
AP3	131	138	6%
AP4	50	49	-1%
AP5	76	82	9%
Σ	415	415	0%

Tendo em vista que não foi observada uma tendência de sábado ter maior ou menor IM em relação ao dia útil, o teste estatístico desencadeado buscará testar a hipótese de que as médias são diferentes, ou seja, um teste bicaudal. Os resultados do teste estão dispostos na Tabela 11.

Tabela 17 - Teste t sábado – dia útil

	<i>dia útil</i>	<i>sábado</i>
Média	82,98	83
Variância	900,677	1108
gl	4	
Stat t	-0,00629	
P(T<=t) bi-caudal	0,995284	
t crítico bi-caudal	2,776445	

Percebe-se que a estatística do teste (Stat t) está dentro da região de não rejeição da hipótese nula ($\pm 2,776$), portanto, a um nível de significância de 5%, não há indícios para se rejeitar a hipótese de que o índice de mobilidade entre dia útil e sábado é diferente. Tal resultado também é mostrado pelo valor-P, que indica uma grande distância da região de rejeição da hipótese nula.

A literatura é pouco clara no que toca a intensidade de mobilidade nos sábados, entretanto, Raux et al. (2017) encontrou que, para pessoas empregadas, a redução no número médio de viagens chega a quatro vezes, para estudantes tal variação chega a 14 vezes, os autores agregaram os resultados de sábado e domingo. Já Susilo e Kitamura

(2005) encontrou que sábado segue um padrão semelhante ao de sexta-feira, entretanto, tais dias fogem do padrão existente em outros dias da semana.

Xianyu et al. (2015) descobriu que em média os respondentes efetuam 2,6 viagens em dias de semana e 2,3 nos sábados. Portanto, na literatura coletada não se obteve resultado em consonância com o obtido por este estudo, tal fato pode ser o indicativo de uma peculiaridade da amostra composta por colaboradores do Mozilla Location Service. Supõe-se que a amostra estudada seja composta por indivíduos que exercem atividade ligadas à profissão também aos sábados.

6.3.2. Comparação entre IM de domingo e dia útil

O gráfico 7 exibe os índices de mobilidade de domingo e dia útil, separados por AP. Nota-se que em todas as APs as barras azuis são maiores que as barras em laranja, o que indica que o IM nos dias úteis é maior que domingo. De fato, ao se analisar a Tabela 12 nota-se que há uma variação negativa em todas as APs pesquisadas, em algumas macrozonas ocorre redução mais acentuada, em outras menos, mas há uma clara tendência de redução do IM na amostra estudada. Pode-se constatar pela tabela que a queda em AP1 é mais acentuada, tal resultado é esperado, visto que esta área se caracteriza como o centro comercial da cidade, que interrompe suas atividades no dia de domingo.

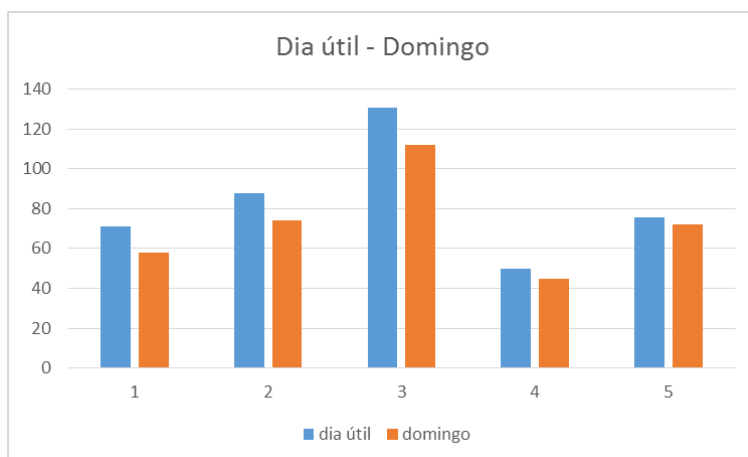


Gráfico 7 - Diferença entre índices de mobilidade - dom

Tabela 18 - Variação de IM entre domingo e dia útil

zona	dia útil	dom	Δ
AP1	71	58	-19%
AP2	88	74	-16%
AP3	131	112	-14%
AP4	50	45	-9%
AP5	76	72	-5%
Σ	415	361	-13%

Por se verificar a tendência de redução do IM com os dias considerados, a hipótese estatística utilizada não será um teste bilateral, pois não se quer testar se as médias entre dias úteis e domingos são apenas diferentes, o objetivo principal do teste é verificar se existem evidências estatísticas de que a média do IM no dia útil é maior que aos domingos.

Tabela 19 - Teste t domingo - dia útil

	<i>dia útil</i>	<i>domingo</i>
Média	82,98	72,2
Variância	900,677	632,2
Observações	5	5
gl	4	
Stat t	3,723134	
P(T<=t) uni-caudal	0,010209	
t crítico uni-caudal	2,131847	

O resultado do teste t é mostrado na tabela 13. Percebe-se que a estatística do teste é mais extrema que o t crítico, portanto, nesse caso, rejeitamos a hipótese de que as médias são iguais. Conclui-se que para um nível de significância de 5% há indícios estatísticos para afirmar que a média do índice de mobilidade do dia útil é maior que a m. Este resultado também é observado pelo valor-P obtido, cabe salientar que a região de não rejeição ficou próxima da estatística do teste calculada, caso fosse escolhido um α de 1%, também comumente utilizado, a hipótese nula seria aceita. O uso de um nível de significância de 0,05 foi utilizado por Hanson e Huff (1988) em seus testes.

Na literatura, muitos autores tratam acerca da intensidade de mobilidade no domingo, e a conclusão é unânime no sentido de que há grande redução em comparação com outros dias da semana, isto é, ainda que atividades sociais sejam realizadas ao domingo, tal como encontrou Allstrom et al. (2016), a tendência é que os indivíduos permaneçam em casa ou realize pequenos deslocamentos.

Susilo e Kitamura (2005), por exemplo, encontraram menor intensidade de mobilidade no domingo em relação aos outros dias da semana, conclusão igualmente obtida por Dharmowijoyo et al. (2015), Raux et al. (2017) e Sagl et al. (2012). Xianyu et al. (2015) descobriu também uma redução no número médio de viagens no final de semana, entretanto, não especificou uma divisão entre sábado e domingo.

6.4. Variabilidade entre dia útil e feriados

Dois feriados foram analisados neste estudo. O dia do trabalho, 1º de Maio, ocorreu no ano de 2017 na segunda feira, já Corpus Christi caiu no dia 15 de Junho, este último por ser um feriado com dia da semana fixo, quinta-feira, comumente é prolongado à sexta-feira.

6.4.1. Comparação entre dia útil e 1º de Maio

A comparação entre o dia do trabalho e o dia útil representativo é dado no gráfico 8. Constatou-se uma diferença considerável entre ambos os dias, bem como uma queda expressiva em AP3. Em AP2, embora em menor grau, também nota-se tal redução.

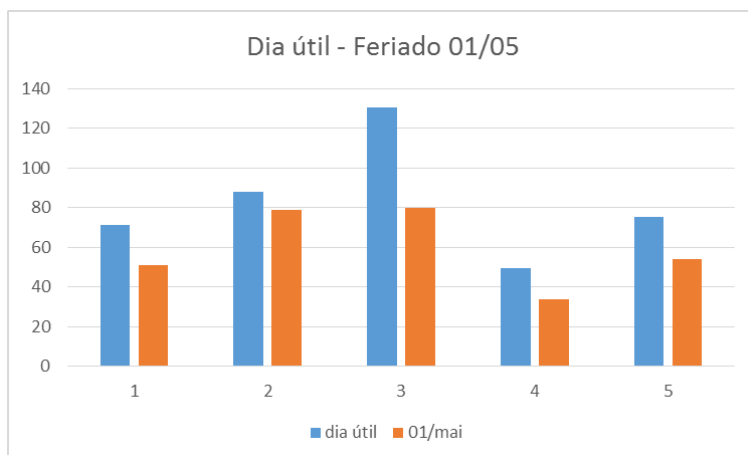


Tabela 20 - Variação de IM entre 01/mai e dia útil

zona	dia útil	01/mai	Δ
AP1	71	51	-28%
AP2	88	79	-10%
AP3	131	80	-39%
AP4	50	34	-32%
AP5	76	54	-28%
Σ	415	298	-28%

Gráfico 8 - Diferença entre índices de mobilidade – 01/mai

A Tabela 14 ressalta as reduções no índice de mobilidade ocorridas em decorrência do dia do trabalho. As maiores quedas são observadas em AP3 e AP4, em AP2 a queda é menor, supõe-se que as amenidades que a macrozona oferece tenha amortecido a queda do IM devido à atração de viagens com o propósito de lazer. Entretanto, a tendência da amostra é de redução do índice em todas as APs. A tabela mostra que provavelmente a média de viagens em dias úteis é maior que nos feriados, um teste de hipótese uni-caudal foi realizado e está disposto na Tabela 15.

Tabela 21 - Teste t 01/mai - dia útil

	<i>dia útil</i>	<i>1-mai</i>
Média	82,98	59,6
Variância	900,677	388,3
Observações	5	5
gl	4	
Stat t	3,255415	
P(T<=t) uni-caudal	0,015608	
t crítico uni-caudal	2,131847	

Com os resultados da Tabela 15 concluiu-se que há evidências estatísticas para afirmar que a média do IM nos dias úteis é maior que no feriado do dia do trabalho. A estatística do teste é mais extrema que o t crítico para a hipótese uni-caudal, assim como

o valor-P é menor que o nível de significância adotado (5%), o que leva a rejeição da hipótese nula, ou seja, a hipótese de que as médias são iguais.

A literatura coletada não contém resultados acerca de intensidade de mobilidade em feriados, apenas resultados acerca de estrutura de decisão relativa a escolha do modo de transporte citada por Yang et al. (2016), no qual em dias úteis os indivíduos priorizam o roteiro ou cronograma, já em feriado prioriza-se o modo de transporte.

6.4.2. Comparação entre dia útil e Corpus Christi (15/jun)

A análise de dois feriados distintos pode trazer conclusões interessantes a esta dissertação, por este motivo optou-se por estudar o feriado de Corpus Christi, assim como o seu prolongamento, o dia 16 de junho, sexta-feira. É possível constatar, pelo gráfico 9, que o IM do dia útil é maior em todas as APs, exceto em AP2, novamente supõe-se que o fato de a zona sul apresentar diversas atrações turísticas e amenidades, pode ter elevado o índice de mobilidade nesta região a ponto de ser um pouco maior que em dias úteis.

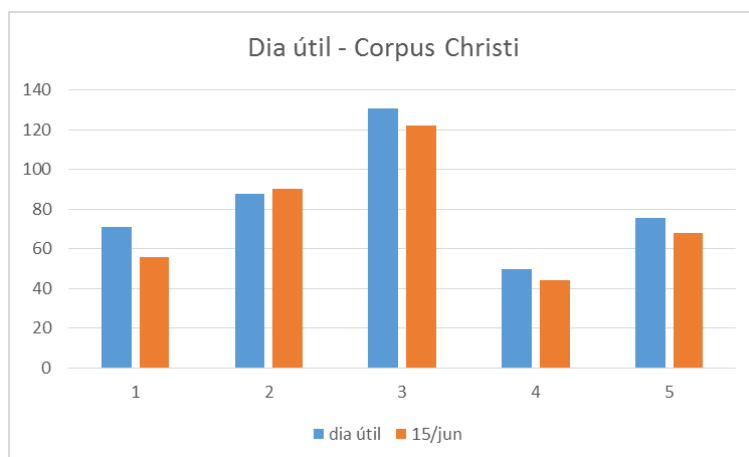


Gráfico 9 - Diferença entre índices de mobilidade – 15/jun

Tabela 22 - Variação de IM entre 15/jun e dia útil

zona	dia útil	15/jun	Δ
AP1	71	56	-21%
AP2	88	90	3%
AP3	131	122	-7%
AP4	50	44	-11%
AP5	76	68	-10%
Σ	415	380	-8%

Na tabela 16 nota-se que a maior variação ocorre na AP1, o que é esperado visto que é a região concentradora de empregos da cidade. Além de AP1, nota-se uma tendência de queda em outras 3 APs, devido a isso optou-se por efetuar um teste unicaudal a fim de testar se a média do dia útil é maior que a média do feriado de Corpus Christi. Os resultados estão dispostos na Tabela 17.

Novamente a estatística do teste superou o valor crítico, levando à rejeição da hipótese nula. Além disso, o valor-P ficou abaixo do nível de significância, o que leva a mesma conclusão, isto é, rejeita-se a hipótese de que as médias são iguais, portanto,

existem indícios de que a média do IM no dia útil é maior que no feriado de Corpus Christi.

Tabela 23 - Teste t Corpus Christi - dia útil

	<i>dia útil</i>	<i>15-jun</i>
Média	82,98	76
Variância	900,677	950
Observações	5	5
gl	4	
Stat t	2,493589	
P(T<=t) uni-caudal	0,033613	
t crítico uni-caudal	2,131847	

Deve-se salientar que a amostra estudada tem a peculiaridade da redução do IM em domingos e feriados, o que não necessariamente é obedecida por todos aqueles que residem na cidade do Rio de Janeiro, entretanto, tal amostra pode ser representativa do conjunto de indivíduos que detém as mesmas características socioeconômicas. Tal explicação é necessária pelo fato de que a queda no IM pode causar estranheza, haja vista a ocorrência de congestionamentos que frequentemente ocorrem nas vias de saída do município em datas referentes a feriados. A caracterização do colaborador MLS pode ser linha de pesquisa para desenvolvimento futuro.

6.4.3. Comparação entre dia útil e 16/jun

A análise do prolongamento do feriado de Corpus Christi foi feito no intuito de verificar o nível com que a mobilidade é afetada em períodos que, na prática, não são feriados, mas que são diretamente afetados por eles. Pelo gráfico 10 observa-se que as APs de 2 a 5 apresentam ligeiras quedas no dia 16 de junho em comparação com o dia útil, já a AP1 apresentou ligeiro aumento.

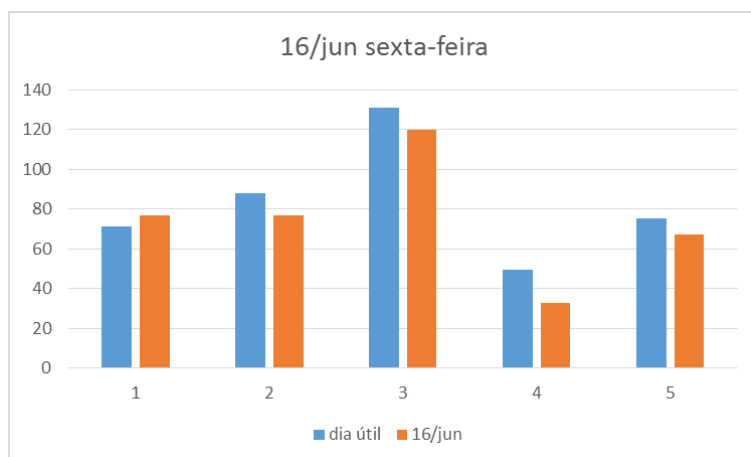


Gráfico 10 - Diferença entre índices de mobilidade – 16/jun

Tabela 24 - Variação de IM entre 16/jun e dia útil

zona	dia útil	16/jun	Δ
AP1	71	77	8%
AP2	88	77	-12%
AP3	131	120	-8%
AP4	50	33	-34%
AP5	76	67	-11%
Σ	415	374	-10%

Numericamente as variações são dadas na Tabela 18, supõe-se, por tais números, que parte da amostra que depende do Centro comercial realizou viagens até esta AP, tais viagens poderiam ter propósito de trabalho ou compras, caso alguns estabelecimentos não tenham prolongado o feriado. Outra hipótese é de que as atrações do centro histórico possam ter atraído viagens neste período, haja vista a presença de vários museus, entre outros atrativos conforme descritos no capítulo anterior. Dentre as duas hipóteses a de maior possibilidade de estar correta é a segunda, uma vez que as viagens a trabalho apresentam baixa variabilidade e maior previsibilidade, já as viagens para lazer tendem a ser mais randômicas (TARIGAN E KITAMURA, 2009), o que pode levar a um aumento não esperado, tal como ocorreu em AP1.

Ainda que haja um aumento no IM na AP1, a tendência observada pela soma das variações é de que o IM médio do dia útil é maior do que o índice no prolongamento do feriado de Corpus Christi. Um teste uni-caudal foi realizado, os resultados estão dispostos na tabela 19.

Tabela 25 - Teste t 16/jun - dia útil

	<i>dia útil</i>	<i>16-jun</i>
Média	82,98	74,8
Variância	900,677	965,2
Observações	5	5
gl	4	
Stat t	2,18095	
P(T<=t) uni-caudal	0,04732	
t crítico uni-caudal	2,13184	

Os resultados tanto da estatística do teste, quanto do valor-P leva a rejeição da hipótese nula, com isso, têm-se evidências de que em média o IM do dia útil é maior que do ocorrido no prolongamento do feriado, embora para AP1 o índice observado tenha sido maior.

6.4.4. Comparação entre feriados

Os feriados analisados ocorreram em datas relativamente próximas, apenas 45 dias de diferença, portanto, não houveram interferências relativas à período festivo ou de férias escolares na mobilidade da amostra, considerando as duas datas estudadas neste trabalho. Dessa maneira, a maior diferença entre os feriados estudados foi o dia da semana em que eles ocorreram. O feriado de Corpus Christi ocorre na quinta-feira com o prolongamento comum de sexta-feira, já o feriado do dia do trabalho, no ano de 2017,

ocorreu numa segunda-feira. Neste subcapítulo pretende-se responder a seguinte pergunta: o dia da semana em que ocorre o feriado interfere no índice de mobilidade?

Em busca de responder o questionamento do parágrafo anterior foi realizado um teste unilateral a fim de verificar a existência de indícios estatísticos de que o IM no feriado de Corpus Christi é maior que o IM no feriado do dia do trabalho. A escolha do teste unilateral foi feita em virtude de os valores absolutos do IM de Corpus Christi serem maiores que os índices do dia do trabalho, conforme disposto na Tabela 20.

Tabela 26 - Variação de IM entre feriados

	15/jun	01/mai	Δ
AP1	56	51	-9%
AP2	90	79	-12%
AP3	122	80	-34%
AP4	44	34	-23%
AP5	68	54	-21%

O resultado do teste t pareado é mostrado na Tabela 21. Os valores mostraram que a hipótese de que as médias são iguais é rejeitada ($p < 0,05$), o que leva a inferência de que o índice de mobilidade no feriado de Corpus Christi é maior.

Tabela 27 - Teste t entre feriados

	15-jun	1-mai
Média	76	59,6
Variância	950	388,3
Observações	5	5
gl	4	
Stat t	2,499233	
P(T<=t) uni-caudal	0,033410	
t crítico uni-caudal	2,131846	

Duas suposições podem ser formuladas diante dos indícios fornecidos pelos dados. A primeira suposição é de que os feriados prolongados estimulam a mobilidade, e a segunda hipótese é de que feriados na segunda-feira estimulam o descanso. Intuitivamente, estas conclusões não parecem estar em desacordo com a realidade, visto que nos feriados de segunda-feira tem-se a obrigatoriedade de exercer atividades cotidianas no outro dia, o que não ocorre nos feriados prolongados de quinta-feira. Com isso, na segunda-feira o indivíduo tende a optar por descansar em vista da semana que, neste caso, começará na terça-feira. É provável que o mesmo raciocínio possa ser

aplicado ao domingo, tendo em vista que este dia tem o menor IM dentre todos os dias da semana.

6.5. Cenário recomendado para uso dos dados MLS

O estudo apresentado nesta dissertação utilizou um tipo de dado com diversas limitações quanto voltados a área de transportes. Além disso, o fato de nenhum estudo anterior ter utilizado tais dados dificultou ainda mais a interpretação e processamento dos mesmos. No entanto, no decorrer do estudo, diversas ideias surgiram no intuito de melhorar a qualidade dos dados, ideias que não foram adotadas neste trabalho devido à restrições de tempo.

Neste subcapítulo busca-se fornecer um cenário possível de uso dos dados MLS, tendo como base toda a experiência adquirida neste ano de trabalho, bem como na literatura coletada. Uma das principais limitações do método utilizado é a não disponibilidade do id do aparelho, o que levou ao estabelecimento das premissas do capítulo 3. Uma maneira de superar esta limitação é a elaboração de um estudo controlado, por meio da colaboração de voluntários. Portanto, o primeiro passo é a obtenção de voluntários, etapa que demandará mais recursos e tempo, a depender do nível de precisão e representatividade que se deseja obter com tal pesquisa.

A segunda etapa é o treinamento dos voluntários. É essencial que os participantes entendam a importância do estudo a fim de que ele se preocupe em reportar os dados da maneira mais fiel possível. Neste treinamento, que poderia ocorrer por vídeo-aula ou apenas cartilhas, será explanado o funcionamento do Mozilla Stumbler. A terceira e última etapa é a coleta de dados. Tal coleta será realizada de duas maneiras, uma única, por e-mail ou entrevista, na qual serão coletados dados socioeconômicos dos voluntários, e outra diária, relativa aos dados locacionais do MLS. Quanto aos dados locacionais é essencial que o indivíduo reporte a hora de upload diariamente, a fim de possibilitar sua triagem na base de dados

Este procedimento tende a melhorar a qualidade dos dados obtidos, sem aumentar em demasiado os custos da pesquisa, um dos principais objetivos dos métodos alternativos. Um fluxograma referente ao método disposto neste subcapítulo é retratado na figura 21.

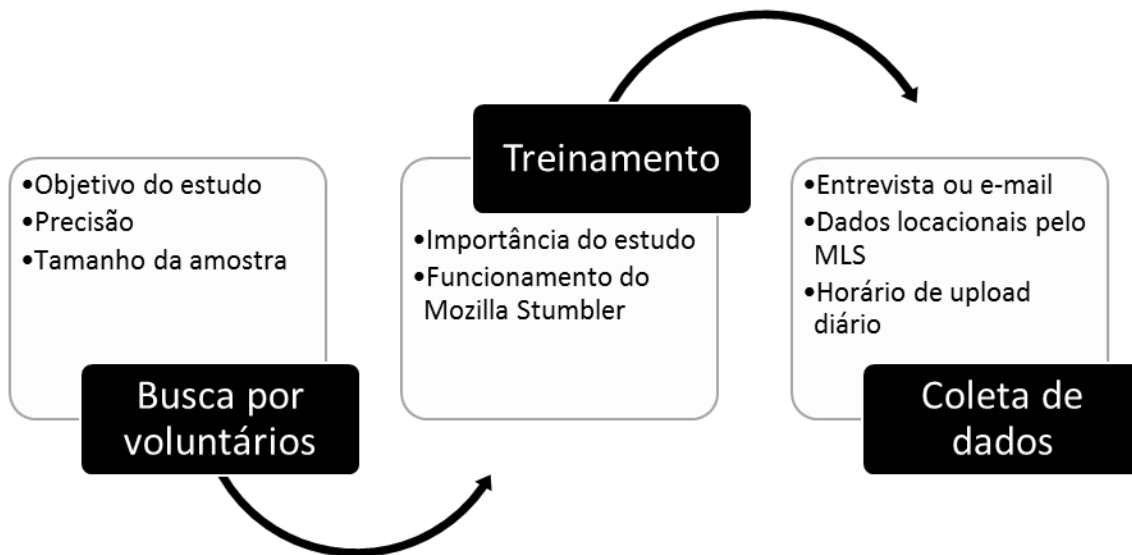


Figura 21 – Método recomendado para uso de dados MLS

Com este método pretende-se coletar dados de maior qualidade e que possam ou confirmar ou refutar as suposições feitas neste trabalho.

7. CONCLUSÃO E RECOMENDAÇÕES

A modelagem em transportes é feita comumente adotando um dia de análise. Tal procedimento considera que os padrões de viagem em outros dias vão obedecer a mesma rotina ou mudar de maneira não significativa. No entanto, as pessoas não tem as mesmas necessidades diárias, tampouco restrições à mobilidade são as mesmas geralmente, fatores que levam a determinada variabilidade no padrão de viagem dos indivíduos.

A análise desta variação de padrão importa diretamente na validade e confiabilidade das pesquisas em transporte que utilizam apenas um dia de pesquisa, tal como o PDTU-Rio que coleta dados de viagens do dia anterior à pesquisa. Portanto, entende-se que o estudo da variabilidade tende a resultar numa maior precisão às pesquisas de origem-destino, tanto é verdade que diversos cidades já investem na coleta de dados longitudinais no intuito de aprimorar seus planos de transporte.

As matrizes origem-destino são geradas a partir de movimentos realizados durante um dia de semana, julgando-os rotineiros, e na maioria das vezes os dados relativos a finais de semana e feriados são desprezados. O estudo destes dias pode trazer conhecimentos no sentido de otimizar a distribuição do transporte público, bem como a mitigação dos efeitos indesejados, como por exemplo, congestionamentos.

Para o estudo da variabilidade existe a possibilidade do uso de métodos tradicionais e alternativos, desde que sejam obtidos dados longitudinais, essenciais à visualização da variação no padrão de viagem. Neste estudo foi utilizado os dados do Mozilla Location Service, serviço que fornece um banco de dados diário contendo, em resumo, informações locais coletadas pelos seus colaboradores. Os dados são de domínio público e, até este estudo, não tinham sido utilizados na área de transportes.

O objetivo deste estudo é analisar a variabilidade da mobilidade entre dia útil, final de semana e feriado por meio de dados do MLS. Para alcançar tal objetivo um método foi elaborado no intuito de extrair informações dos dados obtidos, tal método utilizou softwares de uso recorrente para gerenciamento de base de dados relacional sendo necessário apenas conhecimentos básicos de linguagem SQL.

O índice de mobilidade leva em conta a intensidade com que a rede de transporte de determinada região é solicitada pelos colaboradores do MLS. A variação diária deste índice é um indicativo de demanda de transporte, sendo útil à tomada de decisões, sobretudo, àquelas voltadas ao transporte público. Utilizou-se a análise de variância para análise estatística entre dias úteis e o teste t pareado, com nível de

significância de 5%, para testar médias dos índices entre dia útil, final de semana e feriados.

Os resultados obtidos por este estudo permitem a verificação das hipóteses formuladas, tal verificação é feita abaixo:

1. A variabilidade no dia de semana é significativa, sendo que sexta-feira é o dia de maior índice de mobilidade

Muito embora o total de registros na sexta-feira seja maior, a análise de variância demonstrou que não há diferença significativa entre os grupos. Entende-se, portanto, que há um comportamento habitual e rotineiro no impacto resultante da amostra, o que leva a crer que a análise de apenas um dia de viagens é suficiente para modelar o comportamento do conjunto de indivíduos de maneira confiável.

2. O índice de mobilidade é reduzido igualmente no sábado e domingo em comparação ao dia útil

Embora alguns estudos tenham achado que a redução de mobilidade é igual tanto no sábado quanto no domingo, esta dissertação concluiu que a mobilidade no sábado tem a mesma intensidade do dia útil. Supõe-se que este comportamento se deva ou ao trabalho em meio período ou a um período dedicado a compras. Já domingo parece ser um dia voltado ao descanso, uma vez que o índice de mobilidade é mínimo em comparação com os outros dias da semana.

3. O índice de mobilidade é reduzido igualmente nos feriados de Corpus Christi (15/06) e Dia do Trabalho (01/05) em comparação ao dia útil

O IM é reduzido tanto no Dia do Trabalho quanto em Corpus Christi, entretanto, a redução foi mais significativa no feriado do mês de maio. Supõe-se que o fato de o feriado ocorrer numa segunda-feira impacta diretamente na mobilidade em comparação com feriados prolongados de quinta-feira, nos primeiros tem-se a preferência por permanecer em casa.

Dentre as contribuições deste estudo podem-se citar, primeiramente, o estudo da variabilidade para uma cidade com características distintas das que já foram abordadas pela literatura. Além disso, o uso de uma fonte de dados pública e que ainda não tinham sido utilizadas com o propósito de estudo em transportes. E por último, o estudo agregado dos dados de variabilidade, a fim de saber o impacto na rede de

transporte da resultante de todas as variações, o que não foi realizado em estudos anteriores.

Uma das limitações deste estudo é a não possibilidade de se obter o id de um indivíduo, uma maneira de superar tal limitação se dá por meio da recomendação desenvolvida no capítulo 4.6, portanto, com vistas a estudos futuros, recomenda-se a elaboração de um estudo controlado a fim de possibilitar a obtenção de maior detalhamento de informações. A delimitação geográfica utilizada exigiu a consideração de áreas muito grandes, sugere-se a busca de resultados em níveis mais detalhados, tal como o nível de Região Administrativa ou mesmo os bairros. Além disso, recomenda-se o estudo da variabilidade entre as diferentes estações do ano, sobretudo, a variação existente entre o inverno e verão. Estas sugestões podem aprimorar este estudo e propiciar informações importantes relativas ao planejamento de transporte da cidade do Rio de Janeiro.

REFERÊNCIAS BIBLIOGRÁFICAS

- AHAS, Rein et al. Daily rhythms of suburban commuters' movements in the Tallinn metropolitan area: Case study with mobile positioning data. **Transportation Research Part C: Emerging Technologies**, v. 18, n. 1, p. 45-54, 2010.
- ALLSTRÖM, Andreas; KRISTOFFERSSON, Ida; SUSILO, Yusak. Smartphone Based Travel Diary Collection: Experiences from a Field Trial in Stockholm. In: **European Transport Conference 2016 Association for European Transport (AET)**. 2016.
- ALECRIM, Emerson. Mozilla lança projeto de geolocalização que se baseia em redes Wi-Fi públicas (2013). Disponível em: <<https://tecnoblog.net/143847/mozilla-projeto-de-geolocalizacao-redes-wi-fi-publicas/>>. Acesso em: 06 de outubro de 2017
- ALEKSANDERSEN, Daniel. Help Mozilla build out their Location Service while walking about with your phone (2016). Disponível em: <<https://www.ctrl.blog/entry/contribute-mls-wifi>>. Acesso em: 06 de outubro de 2017
- ALEM, Adriano. Breve relato sobre a formação das divisões administrativas na Cidade do Rio de Janeiro: período de 1961 a 2007. Coleção Estudos Cariocas. Rio de Janeiro: PMRJ, Novembro de 2007.
- AMARAL, Ernesto. Aula 01 Principais Conceitos em Econometria. Métodos Quantitativos de Avaliação de Políticas Públicas. 2010.
- ASAKURA, Yasuo; HATO, Eiji; SUGINO, Katsutoshi. Simulating travel behaviour using location positioning data collected with a mobile phone system. In: **Simulation Approaches in Transportation Analysis**. Springer US, 2005. p. 183-204.
- BEKHOR, Shlomo; COHEN, Yehoshua; SOLOMON, Charles. Evaluating long-distance travel patterns in Israel by tracking cellular phone positions. **Journal of Advanced Transportation**, v. 47, n. 4, p. 435-446, 2013.
- BRITO, Phillippe M. F. Utilização do Sistema de bilhetagem eletrônica para o planejamento operacional em um sistema ferroviário de passageiros. **Dissertação de mestrado**. COPPE/UFRJ, 2016.
- BONNEL, Patrick et al. Passive mobile phone dataset to construct origin-destination matrix: potentials and limitations. **Transportation Research Procedia**, v. 11, p. 381-398, 2015.
- ASSOCIAÇÃO BRASILEIRA DE TELECOMUNICAÇÕES. Mapas das Erbs (Antenas). Disponível em: <<http://www.telebrasil.org.br/panorama-do-setor/mapa-de-erbs-antenas>> Disponível em: 17 de outubro de 2017
- CACERES, N.; WIDEBERG, J. P.; BENITEZ, F. G. Deriving origin-destination data from a mobile phone network. **IET Intelligent Transport Systems**, v. 1, n. 1, p. 15-26, 2007.

- CALABRESE, Francesco et al. Estimating Origin-Destination flows using opportunistically collected mobile phone location data from one million users in Boston Metropolitan Area. **IEEE Pervasive Computing**, v. 99, 2011.
- CALABRESE, Francesco et al. Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. **Transportation research part C: emerging technologies**, v. 26, p. 301-313, 2013.
- DEPARTAMENTO NACIONAL DE INFRAESTRUTURA DE TRANSPORTES. Manual de Estudos de Tráfego. **Publicação IPR – 723**. Rio de Janeiro, 2006.
- DHARMOWIJOYO, Dimas BE; SUSILO, Yusak O.; KARLSTRÖM, Anders. Day-to-day variability in travellers' activity-travel patterns in the Jakarta metropolitan area. **Transportation**, v. 43, n. 4, p. 601-621, 2015.
- DHARMOWIJOYO, Dimas BE et al. Collecting a multi-dimensional three-weeks household time-use and activity diary in the Bandung Metropolitan Area, Indonesia. **Transportation Research Part A: Policy and Practice**, v. 80, p. 231-246, 2015b.
- DHARMOWIJOYO, Dimas BE; SUSILO, Yusak O.; KARLSTRÖM, Anders. Analysing the complexity of day-to-day individual activity-travel patterns using a multidimensional sequence alignment model: A case study in the Bandung Metropolitan Area, Indonesia. **Journal of Transport Geography**, v. 64, p. 1-12, 2017.
- DOS SANTOS MELO, Ângelo et al. Abordagem de baixo custo para coleta de dados de transporte público usando smartphones. **XXIX Congresso Nacional de Pesquisa em Transporte da Anpet**. 2015.
- ELANGO, Vetri; GUENSLER, Randall; OGLE, Jennifer. Day-to-day travel variability in the commute Atlanta, Georgia, study. **Transportation Research Record: Journal of the Transportation Research Board**, n. 2014, p. 39-49, 2007.
- FERIANCIC, Gabriel; CELEIRO, Francisco Raimundo; SILVA, Luiz Norberto Branquinho. Planejamento da Mobilidade com Big Data de Telefonia Móvel. In: **20º Congresso Brasileiro de Transporte e Trânsito, Santos**. 2015.
- FERRAZ, Antonio Clovis Pinto; TORRES, Isaac Guilherme Espinosa. Transporte Público Urbano. 2. ed. São Carlos: Rima, 2004.
- GOOGLE. Política de Privacidade – Informações que coletamos (2017). Disponível em: <<https://www.google.com/intl/pt-BR/policies/privacy/#infocollect>>. Acesso em: 06 de outubro de 2017
- GEIGER, Pedro P. A MetrÓpole e Cidade do Rio de Janeiro. Rio de Janeiro, Portal Geo, 2002. Disponível em: <http://portalgeo.rio.rj.gov.br/bairros Cariocas/texto_cidade.htm>. Acesso em: 10 de outubro de 2017

- GERHARDT, T. e Silveira, D. Métodos de pesquisa. Planejamento e Gestão para o Desenvolvimento Rural da SEAD/UFRGS. – Porto Alegre: Editora da UFRGS, 2009.
- GUERRA, André Leite; BARBOSA, Heloisa Maria; DE OLIVEIRA, Leise Kelli. Estimativa de matriz origem/destino utilizando dados do sistema de bilhetagem eletrônica: proposta metodológica. **TRANSPORTES**, v. 22, n. 3, p. 26-38, 2014.
- HANSON, Susan; HUFF, James O. Assessing day-to-day variability in complex travel patterns. **Transportation Research Record**, v. 891, p. 18-24, 1981.
- HANSON, Susan; HUFF, O. James. Systematic variability in repetitive travel. **Transportation**, v. 15, n. 1, p. 111-135, 1988.
- HUFF, James O.; HANSON, Susan. Repetition and variability in urban travel. **Geographical Analysis**, v. 18, n. 2, p. 97-114, 1986.
- JONES, Peter; CLARKE, Mike. The significance and measurement of variability in travel behaviour. **Transportation**, v. 15, n. 1-2, p. 65-87, 1988.
- KANG, Hejun; SCOTT, Darren M. Exploring day-to-day variability in time use for household members. **Transportation Research Part A: Policy and Practice**, v. 44, n. 8, p. 609-619, 2010.
- LONG, Xuelian; JIN, Lei; JOSHI, James. Exploring trajectory-driven local geographic topics in foursquare. In: **Proceedings of the 2012 ACM Conference on Ubiquitous Computing**. ACM, 2012. p. 927-934.
- MCNEILL, Graham; BRIGHT, Jonathan; HALE, Scott A. Estimating local commuting patterns from geolocated Twitter data. **EPJ Data Science**, v. 6, n. 1, p. 24, 2017.
- MOTAHARI, Sara; ZANG, Hui; REUTHER, Phyllis. The impact of temporal factors on mobility patterns. In: **System Science (HICSS), 2012 45th Hawaii International Conference on**. IEEE, 2012. p. 5659-5668.
- MOZILLA. Data Export (2013). Disponível em:
<https://mozilla.github.io/ichnaea/import_export.html>. Acesso em: 06 de outubro de 2017
- MONTGOMERY, D.C.; RUNGER, G.C. Estatística aplicada e probabilidade para Engenheiros. 4 ed. Rio de Janeiro: LTC, 2009
- NOULAS, Anastasios et al. An empirical study of geographic user activity patterns in foursquare. **ICwSM**, v. 11, p. 70-573, 2011.
- PAS, Eric I.; KOPPELMAN, Frank S. An examination of the determinants of day-to-day variability in individuals' urban travel behavior. **Transportation**, v. 13, n. 2, p. 183-200, 1986.
- PAS, Eric I. Intrapersonal variability and model goodness-of-fit. **Transportation Research Part A: General**, v. 21, n. 6, p. 431-438, 1987.

- PAS, Eric I. Weekly travel-activity behavior. **Transportation**, v. 15, n. 1-2, p. 89-109, 1988.
- PAS, Eric I.; SUNDAR, Subramanian. Intrapersonal variability in daily urban travel behavior: some additional evidence. **Transportation**, v. 22, n. 2, p. 135-150, 1995.
- PREFEITURA DO RIO DE JANEIRO. Anexo técnico I: Informações Sobre Todas as Áreas de Planejamento [2007?]. Disponível em: <<http://www.rio.rj.gov.br/dlstatic/10112/1529762/DLFE-220205.pdf/1.0>>. Acesso em: 10 de outubro de 2017.
- PREFEITURA DO RIO DE JANEIRO. Anexo VI – Descrição e Mapas da área de Planejamento 5 [2011?]. Disponível em: <<http://www.rio.rj.gov.br/dlstatic/10112/4290214/4105682/06.AnexoVIDescricaoeMapadaAreadePlanejamento5.pdf>>. Acesso em: 10 de outubro de 2017.
- PREFEITURA DO RIO DE JANEIRO. População residente por grupos de idade e sexo, segundo as Áreas de Planejamento e Regiões Administrativas - Município do Rio de Janeiro (2011). Disponível em: <http://portalgeo.rio.rj.gov.br/amdpaint_ms.asp?gtema=0&gcod=1>. Acesso em: 12 de outubro de 2017.
- PREFEITURA DO RIO DE JANEIRO. Características geográficas (2009). Disponível em: <<http://www.rio.rj.gov.br/web/riotur/caracteristicas-geograficas>>. Acesso em: 12 de outubro de 2017.
- PREFEITURA DO RIO DE JANEIRO. Gerências de Planejamento Local (2015). Disponível em: <http://www.rio.rj.gov.br/web/smu/exibeconteudo?id=4481419>. Acesso em: 16 de outubro de 2017.
- PREFEITURA DO RIO DE JANEIRO. Relatório 4 – Planejamento e Execução das Pesquisas: Parte 1. **Plano Diretor de Transportes Urbanos do Rio de Janeiro**. 2016
- QUEIROZ, Frank; RIBEIRO, Paulo. Análise do padrão de mobilidade por meio de dados mozilla location service. **XV Rio de Transportes**. 2017
- RAUX, Charles; MA, Tai-Yu; CORNELIS, Eric. Variability in daily activity-travel patterns: the case of a one-week travel diary. **European transport research review**, v. 8, n. 4, p. 26, 2016.
- SANTOS, Ricardo di L. Redes GSM, GPRS, EDGE e UMTS (2008). Disponível em: https://www.gta.ufrj.br/ensino/eel879/trabalhos_vf_2008_2/ricardo/1_1.html Acesso em: 24 de outubro de 2017
- RAIMOND, Ana-Maria Olteanu; COURONNÉ, Thomas. Modeling Humain Behavior in Space and Time Using Mobile Phone Data. In: **Theories and Simulations of Complex Social Systems**. Springer Berlin Heidelberg, 2014. p. 31-42.
- RICHARDSON, Anthony J.; AMPT, Elizabeth S.; MEYBURG, Arnim H. **Survey methods for transport planning**. Melbourne: Eucalyptus Press, 1995.

- SAGL, Günther; LOIDL, Martin; BEINAT, Euro. A visual analytics approach for extracting spatio-temporal urban mobility information from mobile network traffic. **ISPRS International Journal of Geo-Information**, v. 1, n. 3, p. 256-271, 2012.
- SCHLICH, Robert; AXHAUSEN, Kay W. Habitual travel behaviour: evidence from a six-week travel diary. **Transportation**, v. 30, n. 1, p. 13-36, 2003.
- SILVEIRA, D. T.; CÓRDOVA, F. P. Métodos de pesquisa/[organizado por] Tatiana Engel Gerhardt e Denise Tolfo Silveira; coordenado pela Universidade Aberta do Brasil–UAB/UFRGS e pelo Curso de Graduação Tecnológica–Planejamento e Gestão para o Desenvolvimento Rural da SEAD/UFRGS. **Porto Alegre: Editora da UFRGS**, p. 31-32, 2009.
- SPARKS, Kevin A. et al. Facility detection and popularity assessment from text classification of social media and crowdsourced data. In: **Proceedings of the 10th Workshop on Geographic Information Retrieval**. ACM, 2016. p. 2.
- SUSILO, Yusak O.; AXHAUSEN, Kay W. Repetitions in individual daily activity–travel–location patterns: a study using the Herfindahl–Hirschman Index. **Transportation**, v. 41, n. 5, p. 995-1011, 2014.
- SUSILO, Yusak; KITAMURA, Ryuichi. Analysis of day-to-day variability in an individual's action space: exploration of 6-week Mobidrive travel diary data. **Transportation Research Record: Journal of the Transportation Research Board**, n. 1902, p. 124-133, 2005.
- STOPHER, Peter; ZHANG, Yun. Repetitiveness of daily travel. **Transportation Research Record: Journal of the Transportation Research Board**, n. 2230, p. 75-84, 2011.
- TARIGAN, Ari; KITAMURA, Ryuichi. Week-to-week leisure trip frequency and its variability. **Transportation Research Record: Journal of the Transportation Research Board**, n. 2135, p. 43-51, 2009.
- TENUTA, Livia. A Escolha do método estatístico. Universidade Estadual de Campinas. Faculdade de Odontologia de Piracicaba. 2009.
- XIANYU, Jianchuan; RASOULI, Soora; TIMMERMANS, Harry. Analysis of variability in multi-day GPS imputed activity-travel diaries using multi-dimensional sequence alignment and panel effects regression models. **Transportation**, v. 44, n. 3, p. 533-553, 2017.
- YANG, Liya; SHEN, Qing; LI, Zhibin. Comparing travel mode and trip chain choices between holidays and weekdays. **Transportation Research Part A: Policy and Practice**, v. 91, p. 273-285, 2016.
- YU, Chang; HE, Zhao-Cheng. Analysing the spatial-temporal characteristics of bus travel demand using the heat map. **Journal of Transport Geography**, v. 58, p. 247-255, 2017.

WANG, Ming-Heng et al. Estimating dynamic origin-destination data and travel demand using cell phone network data. **International Journal of Intelligent Transportation Systems Research**, v. 11, n. 2, p. 76-86, 2013.

ZHONG, Chen et al. Measuring variability of mobility patterns from multiday smart-card data. **Journal of Computational Science**, v. 9, p. 125-130, 2015.

ANEXO I – BAIRROS POR AP DO RIO DE JANEIRO (Parte I)

AP	Bairros	
1	Benfica	
	Caju	
	Catumbi	
	Centro	
	Cidade Nova	
	Estácio	
	Gamboa	
	Mangueira	
	Paquetá	
	Rio Comprido	
	Santa Teresa	
	Santo Cristo	
	São Cristóvão	
	Saúde	
	Vasco da Gama	
	2	Alto da B. Vista
		Andaraí
Botafogo		
Catete		
Copacabana		
Cosme Velho		
Flamengo		
Gávea		
Glória		
Grajaú		
Humaitá		
Ipanema		
Jardim Botânico		
Lagoa		
Laranjeiras		
Leblon		
Leme		
Maracanã		
Praça da Bandeira		
Rocinha		
São Conrado		
Tijuca		
Urca		
Vidigal		
Vila Isabel		

AP	Bairros
3	Abolição
	Acari
	Água Santa
	Anchieta
	Bancários
	Barros Filho
	Bento Ribeiro
	Bonsucesso
	Brás de Pina
	Cachambi
	Cacuaia
	Campinho
	Cascadura
	Cavalcanti
	Cidade Universit.
	Cocotá
	Coelho Neto
	Colégio
	Comp. do Alemão
	Cordovil
	Costa Barros
	Del Castilho
	Encantado
	Engenheiro Leal
	Engenho da Rainha
	Engenho de Dentro
	Engenho Novo
	Freguesia
	Galeão
	Guadalupe
	Higienópolis
	Honório Gurgel
	Inhaúma
Irajá	
Jacaré	
Jacarezinho	
Jardim América	
Jardim Carioca	
Jardim Guanabara	
Lins de Vasconc.	

ANEXO I – BAIRROS POR AP DO RIO DE JANEIRO (Parte II)

AP	Bairros
3	Madureira
	Manguinhos
	Maré
	Marechal Hermes
	Maria da Graça
	Méier
	Moneró
	Olaria
	Oswaldo Cruz
	Parada de Lucas
	Parque Anchieta
	Parque Colúmbia
	Pavuna
	Penha
	Penha Circular
	Piedade
	Pilares
	Pitangueiras
	Portuguesa
	Praia da Bandeira
	Quintino Bocaiúva
	Riachuelo
	Ribeira
	Ricardo de Alb.
	Rocha
	Rocha Miranda
	Sampaio
	São Fr. Xavier
	Vaz Lobo
	Vicente de Carv.
	Vigário Geral
	Vila da Penha
Vila Kosmos	
Vista Alegre	
Zumbi	
4	Anil
	Barra da Tijuca
	Camorim

AP	Bairros	
4	Cidade de Deus	
	Curicica	
	Freguesia	
	Gardênia Azul	
	Grumari	
	Itanhangá	
	Jacarépaguá	
	Joá	
	Pechincha	
	Praça Seca	
	Recreio dos Band.	
	Tanque	
	Taquara	
	Vargem Grande	
	5	Vargem Pequena
		Vila Valqueire
Bangu		
B. de Guaratiba		
Campo dos Afo.		
Campo Grande		
Cosmos		
Deodoro		
Gericinó		
Guaratiba		
Inhoaíba		
Jardim Sulacap		
Magalhães Bastos		
Paciência		
Padre Miguel		
Pedra de Guarat.		
Realengo		
Santa Cruz		
Santíssimo		
Senador Camará		
Senador Vasc.		
Sepetiba		
Vila Militar		

ANEXO II – OBTENÇÃO DOS DIAS REPRESENTATIVOS

	AP	07/mai	14/mai	21/mai	28/mai	04/jun	11/jun	18/jun	25/jun	02/jul	09/jul	16/jul	23/jul	30/jul	mediana
domingo	1	69	61	71	62	60	56	56	32	46	58	56	57	62	58
	2	101	64	76	65	74	70	71	84	61	95	100	78	66	74
	3	105	92	119	104	105	112	133	118	106	125	121	111	131	112
	4	66	38	63	50	40	32	58	47	39	45	32	52	43	45
	5	68	72	87	51	80	65	81	73	65	64	63	93	74	72
		Feriado	15/mai	22/mai	29/mai	05/jun	12/jun	19/jun	26/jun	03/jul	10/jul	17/jul	24/jul	31/jul	mediana
segunda	1	51	50	74	63	78	56	69	73	70	81	53	82	59	69
	2	79	86	80	76	83	74	79	100	69	106	84	97	76	80
	3	80	93	151	129	144	114	125	131	113	174	129	163	143	129
	4	34	54	43	53	61	55	58	51	43	65	32	46	39	51
	5	54	79	80	77	70	80	83	90	79	67	62	61	89	79
		02/mai	09/mai	30/mai	06/jun	13/jun	20/jun	27/jun	04/jul	11/jul	18/jul	25/jul			mediana
terça	1	66	71	56	63	59	73	105	79	100	69	70			70
	2	91	65	77	76	86	85	87	101	111	89	96			87
	3	119	96	144	122	123	144	138	121	176	129	161			129
	4	65	24	58	72	49	61	56	57	62	34	36			57
	5	86	66	51	81	85	69	70	87	90	63	74			74
		03/mai	10/mai	24/mai	31/mai	07/jun	14/jun	21/jun	28/jun	05/jul	12/jul	19/jul	26/jul		mediana
quarta	1	87	76	64	81	67	67	78	58	54	61	101	78		72
	2	89	100	93	81	68	68	94	90	70	110	87	102		90
	3	158	170	128	149	134	134	131	127	122	128	170	162		134
	4	50	56	48	60	46	46	67	50	49	38	38	38		49
	5	73	83	75	61	66	66	75	65	67	75	85	86		74
		04/mai	11/mai	25/mai	01/jun	08/jun	Feriado	22/jun	29/jun	06/jul	13/jul	20/jul	27/jul		mediana
quinta	1	80	62	61	86	89	56	67	70	59	64	78	102		69
	2	93	79	69	85	80	90	115	116	97	98	81	98		92
	3	132	112	128	123	143	122	100	129	128	122	155	150		128
	4	45	48	53	38	49	44	54	63	59	37	36	45		47
	5	60	68	84	65	78	68	87	68	88	80	81	73		76
		05/mai	12/mai	26/mai	02/jun	09/jun	Pós-Fe	23/jun	30/jun	07/jul	14/jul	21/jul	28/jul		mediana
sexta	1	87	99	69	59	79	77	69	69	70	91	77	82		77
	2	76	91	79	74	99	77	89	101	91	100	104	106		91
	3	105	129	143	113	136	120	151	130	131	144	167	149		134
	4	32	24	70	53	69	33	68	46	54	38	42	45		46
	5	51	47	84	77	64	67	70	73	83	94	94	99		75
		06/mai	13/mai	20/mai	27/mai	03/jun	10/jun	17/jun	24/jun	01/jul	08/jul	15/jul	22/jul	29/jul	mediana
sábado	1	81	54	68	38	60	65	45	68	66	98	79	71	75	68
	2	84	65	52	76	78	98	78	95	74	102	85	98	78	78
	3	112	87	114	137	138	146	122	99	149	143	160	142	171	138
	4	35	48	46	67	49	69	49	49	51	60	70	64	42	49
	5	67	72	65	71	78	83	74	82	82	87	106	113	114	82

Obs: Os feriados, bem como as emendas não são contabilizados no cálculo da mediana

ANEXO III – CÓDIGO PARA AGRUPAMENTO DE REGISTROS

- Criação da tabela com dados agrupados

```
create table agrupados
select updated, hora, `data`, `local` from mls
group by updated, hora, `data`, `local`
order by updated, `local`;
```

- criação da tabela para o dia da semana e realização das contagens das detecções nas zonas

```
create table domingos
select * from agrupados
where `data` = '07/05/2017' or `data` = '14/05/2017'
or `data` = '21/05/2017' or `data` = '28/05/2017'
or `data` = '04/06/2017' or `data` = '11/06/2017'
or `data` = '18/06/2017' or `data` = '25/06/2017'
or `data` = '02/07/2017' or `data` = '09/07/2017'
or `data` = '14/07/2017' or `data` = '23/07/2017'
or `data` = '30/07/2017';
```

```
select `data`, `local`, count(*) from domingos
group by `local`, `data`
order by `data`, `local`;
```