



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA DE TRANSPORTES
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE TRANSPORTES

RENATO GOERSCH ANDRADE PARENTE

**CARACTERIZAÇÃO DOS PADRÕES DE MOBILIDADE DOS USUÁRIOS DE
TRANSPORTE PÚBLICO POR ÔNIBUS EM FORTALEZA ATRAVÉS DO *BIG DATA*
DE BILHETAGEM ELETRÔNICA**

FORTALEZA

2021

RENATO GOERSCH ANDRADE PARENTE

CARACTERIZAÇÃO DOS PADRÕES DE MOBILIDADE DOS USUÁRIOS DE
TRANSPORTE PÚBLICO POR ÔNIBUS EM FORTALEZA ATRAVÉS DO *BIG DATA* DE
BILHETAGEM ELETRÔNICA

Dissertação de Mestrado apresentada à coordenação do Programa de Pós-Graduação em Engenharia de Transportes da Universidade Federal do Ceará como requisito parcial para obtenção do título de Mestre em Engenharia de Transportes. Área de Concentração: Planejamento e Operação de Sistemas de Transportes.

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

FORTALEZA

2021

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- P252c Parente, Renato Goersch Andrade.
Caracterização dos padrões de mobilidade dos usuários de transporte público por ônibus em Fortaleza através do Big Data de bilhetagem eletrônica / Renato Goersch Andrade Parente. – 2021.
147 f. : il. color.
- Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia de Transportes, Fortaleza, 2021.
Orientação: Prof. Dr. Francisco Moraes de Oliveira Neto.
1. Transporte público. 2. Padrões de mobilidade. 3. Análise da demanda. 4. Clusterização. 5. Caracterização da mobilidade. I. Título.

CDD 388

RENATO GOERSCH ANDRADE PARENTE

CARACTERIZAÇÃO DOS PADRÕES DE MOBILIDADE DOS USUÁRIOS DE
TRANSPORTE PÚBLICO POR ÔNIBUS EM FORTALEZA ATRAVÉS DO *BIG DATA* DE
BILHETAGEM ELETRÔNICA

Dissertação de Mestrado apresentada à
coordenação do Programa de Pós-Graduação em
Engenharia de Transportes da Universidade
Federal do Ceará como requisito parcial para
obtenção do título de Mestre em Engenharia de
Transportes. Área de Concentração: Planejamento
e Operação de Sistemas de Transportes.

Aprovada em 30/03/2021.

BANCA EXAMINADORA

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

Prof. Carlos Felipe Grangeiro, Ph.D. (Examinador Interno)
Universidade Federal do Ceará (UFC)

Dr.^a Camila Soares Henrique Fontenele Garcia, Ph.D. (Examinador Externo)
Instituto Superior Técnico – Universidade de Lisboa (ULisboa)

Prof. Francelino Franco Leite de Matos Sousa, Ph.D. (Examinador Externo)
Universidade Christus (Unichristus)

AGRADECIMENTOS

Primeiramente, gostaria de agradecer a minha esposa e companheira, Thais. Você é peça fundamental na motivação de tudo que faço nessa vida. Me apoiou e esteve ao meu lado durante toda essa caminhada do mestrado. Obrigado por ser o meu norte, sempre. Obrigado pela paciência quando foi preciso. Obrigado por estar sempre ao meu lado, não importa o que aconteça. Aproveito também, para agradecer ao meu cachorro, Mogli. Que esteve ao meu lado durante horas e horas na realização deste trabalho. Que em muitas de nossas caminhadas, me possibilitou momentos de reflexão sobre as diversas situações da vida. Vocês dois são tudo de mais precioso que tenho nessa vida.

Aos meus pais, Francisco e Gardênia. Que se dedicaram ao máximo para me proporcionar a melhor educação possível, mesmo nos tempos mais difíceis. Sem vocês, nada disso seria possível. Minha mãe, em especial, por ter plantado em mim a vontade de realizar um mestrado. Meu pai, que me molda diariamente com a sua sensibilidade com as pessoas e o mundo ao seu redor. Muito obrigado!

Ao professor Moraes, por ter acreditado no meu potencial e do meu trabalho, e por ter me incentivado nos momentos mais difíceis do mestrado. Obrigado, pelas inúmeras horas que estivemos em reuniões nas sextas-feiras à tarde. Suas sugestões e *insights* foram essenciais para a produção deste trabalho. Serei eternamente grato, e espero que ainda possamos compartilhar outras experiências.

Ao professor Felipe, o qual tenho grande admiração pela excepcional pessoa e docente que és. Obrigado por ter me guiado durante o primeiro ano e as primeiras disciplinas do mestrado, elas foram essenciais para a minha formação pessoal e intelectual.

À Camila, que fez parte da minha banca de qualificação e agora faz parte na banca defesa. Obrigado por sua participação, seus comentários e *feedbacks*, assim como sua pesquisa de mestrado e tese de doutorado, foram importantíssimos para o desenvolvimento deste trabalho.

Ao Programa de Pós-Graduação em Engenharia de Transportes da UFC (PETRAN) e os seus professores, que me permitiram vivenciar um ambiente acadêmico incrível a qual eu não conhecia.

Ao André, o qual tive o prazer de conviver diariamente durante os dois anos que trabalhei na Gerência de Transportes e Integração do METROFOR. Obrigado por ter me

apresentado e incentivado a ingressar no PETRAN e por todo conhecimento compartilhado durante os dois anos.

Aos colegas de mestrado e alguns alunos de doutorado, Amelia, Artur, Bianca, Davi, Diego, Leonardo, Nara, Nauri, Pedro e Wendy. Que compartilharam comigo os diversos momentos proporcionados pelo PETRAN. Obrigado também, em especial, ao João Lucas, que esteve sempre disponível para ajudar no desenvolvimento deste trabalho.

Agradeço também à Secretaria Municipal de Conservação e Serviços Públicos (SCSP) da Prefeitura de Fortaleza, pela disponibilização dos dados utilizados nesta pesquisa.

Ao CNPq, por ter financiado esta pesquisa.

RESUMO

O sistema de transporte público no Brasil vem perdendo demanda nas últimas décadas. Devido ao importante papel desse sistema em promover a equidade no acesso, é essencial, para o planejamento de transportes, conhecer os padrões de mobilidade dos usuários e como estes padrões variam espacialmente e temporalmente. Em Fortaleza, mesmo com o investimento em políticas de transporte que buscam melhorar as experiências dos usuários, esta queda também é percebida. Contudo, não se sabe quais são os usuários responsáveis por esta queda, ou quais as características desses usuários, quais regiões eles residem ou quais grupos sociais eles representam. Acredita-se que vasta disponibilidade de dados coletados continuamente pelo sistema de *Smart Card* dos ônibus seja um caminho alternativo para a caracterização da mobilidade dos usuários do sistema. Porém, devido ao grande volume, há complexidades em transformar esses dados em informações úteis para o planejamento de transportes. Esta pesquisa utiliza ferramentas de clusterização para identificar e caracterizar padrões de mobilidade dos usuários no *Big Data* de transporte público por ônibus. A partir da revisão de literatura, algumas hipóteses sobre os padrões de mobilidade são levantadas e sobre como esses padrões se relacionam com a queda na demanda e com problemas de mobilidade. A partir das hipóteses, análises exploratórias, descritivas e espaciais de indicadores de mobilidade foram realizadas. Essas análises permitiram a identificação de padrões de mobilidade para os usuários do sistema e ajudaram a compreender como esses padrões vem alterando espacialmente e temporalmente.

Palavras-chave: transporte público; *big data*; *smart card*; análise da demanda; padrões de mobilidade; clusterização; caracterização da mobilidade.

ABSTRACT

The public transport system in Brazil has been losing demand in recent decades. Due to the important role of this system in promoting equity in access, it is essential, for transport planning, to know the mobility patterns of users and how these patterns vary spatially and temporally. In Fortaleza, even with the investment in transport policies that seek to improve the experiences of users, this drop is also perceived. However, it is not known which users are responsible for this fall, or what the characteristics of these users are, which regions they reside in or which social groups they represent. It is believed that the vast availability of data collected continuously by the Smart Card system of buses is an alternative way to characterize the mobility of users of the system. However, due to its large volume, there are complexities in transforming this data into useful information for transportation planning. This research uses clustering tools to identify and characterize user mobility patterns using Big Data generated by bus public transport system. Guided by the literature review, some hypotheses about mobility patterns are raised and about how these patterns are related to falling demand and mobility problems. Based on the hypotheses, exploratory, descriptive, and spatial analyzes of mobility indicators were carried out. These analyzes allowed the identification of mobility patterns for users of the system and helped to understand how these patterns have been changing spatially and temporally.

Keywords: public transportation; big data; smart card; demand analysis; mobility patterns; clustering; mobility characterization.

LISTA DE ILUSTRAÇÕES

Figura 1 – Definições de Acessibilidade e Mobilidade.....	21
Figura 2 – Classificação dos problemas de acessibilidade e mobilidade	24
Figura 3 – Relações causais por categorias de restrições de acessibilidade.....	26
Figura 4 – Comparação entre clusters	47
Figura 5 – Representação do DBSCAN	50
Figura 6 – Representação do <i>k-means</i>	52
Figura 7 – Representação dos resultados de uma clusterização HAC.....	54
Figura 8 – Mapa de quintis de densidade populacional da população de baixa renda e alta renda.....	62
Figura 9 – Mapa de quintis da densidade de empregos da população de baixa renda e alta renda	63
Figura 10 – Distribuição espacial das matrículas para o ano de 2015.....	64
Figura 11 – Oferta do sistema de transporte.....	65
Figura 12 – Rede de transporte público de Fortaleza	66
Figura 13 – Mapa de Quintis da densidade de frequência de linhas e pontos de parada	68
Figura 14 – Distribuição espacial da acessibilidade da baixa renda para 2015.....	69
Figura 15 – Proposta metodológica de caracterização dos padrões de mobilidade.....	70
Figura 16 – Representação da problemática.....	71
Figura 17 – Método de definição do número de padrões esperados e pesos dos atributos	83
Figura 18 – Metodologia de Caracterização dos Padrões de Mobilidade	85
Figura 19 – Proposta de Zoneamento para Fortaleza	89
Figura 20 – Distribuição de frequência relativa dos atributos.....	94
Figura 21 – Matriz de correlação entre os atributos	95
Figura 22 – Método do cotovelo (Elbow Method).....	96
Figura 23 – Distribuição dos usuários por padrão para N=3, 4 e 5	97
Figura 24 – Distribuições das frequências relativas dos atributos para cada padrão (pesos iguais)	98
Figura 25 – Distribuições das frequências relativas dos atributos para cada padrão (pesos 2 para distância temporal média e número de dias)	100

Figura 26 – Proporções por tipo de cartão de cada grupo (2014 e 2018), e diferença de usuários entre os anos 2018 e 2014.....	101
Figura 27 – Distribuições das distâncias dos clusters de validações aos endereços de residência e distâncias do clusters de validações a parada mais próxima das residências.....	102
Figura 28 – Mapa de quintis; Lisa Map e Boxplot para o indicador proporções dos padrões no ano de 2018.....	104
Figura 29 – Mapa de quintis; Lisa Map e Boxplot para o indicador de frequência média no ano de 2018	106
Figura 30 – Mapa de quintis; Lisa Map e Boxplot para o indicador de número de dias típicos de uso no ano de 2018.....	107
Figura 31 – Mapa de quintis; Lisa Map e Boxplot para a viagens produzidas em um dia útil por zona no ano de 2018.....	108
Figura 32 – Mapa de quintis; Lisa Map e Boxplot para os domicílios de baixa renda (2018)...	109
Figura 33 – Mapa de quintis; Lisa Map e Boxplot para o indicador viagens por domicílios (2018)	109
Figura 34 – Mapa de quintis; Lisa Map e Boxplot para o indicador viagens por domicílios (2018)	110
Figura 35 – Diagrama de Sankey do fluxo de usuários entre os padrões de 2014 para 2018	112
Figura 36 – Mapa de quintis; Lisa Map e Boxplot para o indicador diferença de viagens entre 2018 e 2014.....	113
Figura 37 – Mapa de quintis; Lisa Map e Boxplot para o indicador diferença de viagens	115
Figura 38 – Mapa de quintis; Lisa Map e Boxplot para o indicador diferença de viagens por domicílios.....	116
Figura 39 – Mapa de quintis; Lisa Map e Boxplot para o indicador proporção da diferença das viagens por domicílios pelo o valor das viagens por domicílios no ano de 2014...	117
Figura 40 – Mapa de quintis; Lisa Map e Boxplot para o indicador proporção da diferença das viagens por domicílios pelo o valor das viagens por domicílios de 2014 para cada padrão	118
Figura 41 – Mapa de quintis; Lisa Map e Boxplot para o indicador de frequência média por padrão para o ano de 2014.....	133

Figura 42 – Mapa de quintis; Lisa Map e Boxplot para o indicador de número de dias de uso por padrão para o ano de 2014	134
Figura 43 – Mapa de quintis; Lisa Map e Boxplot para o indicador de proporções dos padrões para o ano de 2014.....	135
Figura 44 – Mapa de quintis; Lisa Map e Boxplot para o indicador viagens produzidas em um dia útil por padrão para o ano de 2014.....	136
Figura 45 – Mapa de quintis; Lisa Map e Boxplot para o indicador quantidade de usuários por zona para o ano de 2014	137
Figura 46 – Mapa de quintis; Lisa Map e Boxplot para o indicador quantidade de viagens produzidas por zona para o ano de 2014.....	137
Figura 47 – Mapa de quintis; Lisa Map e Boxplot para o indicador quantidade de viagens por domicílios por zona para o ano de 2014.....	138
Figura 48 – Mapa de quintis; Lisa Map e Boxplot para o indicador quantidade de usuários por padrão para o ano de 2014.....	139
Figura 49 – Mapa de quintis; Lisa Map e Boxplot para o indicador viagens por domicílios por padrão para o ano de 2014.....	140
Figura 50 – Mapa de quintis; Lisa Map e Boxplot para o indicador quantidade de usuários zona para o ano de 2014.....	141
Figura 51 – Mapa de quintis; Lisa Map e Boxplot para o indicador quantidade de usuários por padrão para o ano de 2018.....	142
Figura 52 – Mapa de quintis; Lisa Map e Boxplot para o indicador viagens produzidas em um dia útil por padrão para o ano de 2018	143
Figura 53 – Mapa de quintis; Lisa Map e Boxplot para o indicador diferença de usuários por padrão	144
Figura 54 – Mapa de quintis; Lisa Map e Boxplot para o indicador diferença de viagens por domicílios por padrão.....	145
Figura 55 – Mapa de quintis; Lisa Map e Boxplot para o indicador diferença de usuários por domicílios	146

LISTA DE TABELAS

Tabela 1 – Atributos encontrados na literatura de clusterização a partir de dados de Smart Card	57
Tabela 2 – Lista de hipóteses sobre os padrões de mobilidade.....	72
Tabela 3 – Amostra dados de bilhetagem.	75
Tabela 4 – Amostra dados de cadastro bilhete único.....	77
Tabela 5 – Lista de indicadores	86
Tabela 6 – Medidas de tendência central e variação.....	93
Tabela 7 – Análise descritiva dos clusters	99

SUMÁRIO

1 INTRODUÇÃO	14
1.1 Problemática	14
1.2 Objetivos	18
1.3 Estrutura da dissertação	18
2 CARACTERIZAÇÃO DA MOBILIDADE	20
2.1 Conceitos e práticas (mobilidade x acessibilidade)	20
2.2 Problemas de mobilidade	22
2.3 Características que impactam na acessibilidade e na mobilidade	25
2.4 Indicadores de mobilidade	28
2.5 O papel do transporte público no planejamento da mobilidade	30
2.6 Caracterização da mobilidade em Fortaleza	31
2.7 Padrão de mobilidade	34
2.8 Conclusão e lacunas	35
3 CARACTERIZAÇÃO E IDENTIFICAÇÃO DE PADRÕES DE MOBILIDADE A PARTIR DO <i>BIG DATA</i> DE TRANSPORTE PÚBLICO	39
3.1 <i>Big Data</i> de transporte público	39
3.2 Identificação de padrões de mobilidade a partir de dados de <i>Smart Card</i>	44
3.3 Ferramentas de clusterização (<i>Clustering</i>)	48
3.3.1 <i>Density-based spatial clustering of applications with noise (DBSCAN)</i>	49
3.3.2 <i>K-means</i>	51
3.3.2.1 <i>Número de Grupos (k-means)</i>	52
3.3.3 <i>Hierarchical Ascending Clustering (HAC)</i>	53
3.3.4 <i>Distância de similaridade</i>	54
3.3.5 <i>Atributos (parâmetros de entrada)</i>	56

3.4 Conclusões e lacunas.....	59
4 CONTEXTUALIZAÇÃO DOS SISTEMAS ANALISADOS.....	61
4.1 Uso do solo e atividades.....	62
4.2 Sistemas de transportes.....	64
4.2.1 <i>Sistema de transporte público por ônibus de Fortaleza.....</i>	<i>65</i>
4.3 Problemas de mobilidade para a população de baixa renda.....	68
5 PROPOSTA METODOLÓGICA PARA IDENTIFICAÇÃO E CARACTERIZAÇÃO DOS PADRÕES DE MOBILIDADE A PARTIR DE DADOS DE SMART CARD.....	70
5.1 Problemática.....	71
5.2 Dados.....	73
5.2.1 <i>Dados de Smart Card e GPS.....</i>	<i>73</i>
5.2.2 <i>Dados de GTFS.....</i>	<i>75</i>
5.2.3 <i>Dados de cadastro do bilhete único.....</i>	<i>76</i>
5.3 Identificação dos padrões de mobilidade.....	78
5.3.1 <i>Construção de perfis dos usuários.....</i>	<i>80</i>
5.3.2 <i>Definição da quantidade de grupos e pesos dos atributos.....</i>	<i>82</i>
5.4 Caracterização dos padrões de mobilidade.....	84
5.4.1 <i>Indicadores.....</i>	<i>85</i>
5.4.2 <i>Zonamento.....</i>	<i>88</i>
5.4.3 <i>Análise descritiva e espacial dos padrões.....</i>	<i>90</i>
5.4.4 <i>Análise da evolução temporal.....</i>	<i>91</i>
5.4.5 <i>Análise da evolução espacial.....</i>	<i>91</i>
6 ANÁLISES E DISCUSSÕES DOS RESULTADOS.....	92
6.1 Identificação dos padrões de mobilidade.....	92
6.1.1 <i>Análise descritiva dos atributos.....</i>	<i>92</i>

<i>6.1.2 Análise de correlação dos atributos</i>	94
<i>6.1.2 Aplicação de método estatístico para definição de grupos</i>	95
<i>6.1.4 Definição dos pesos</i>	97
6.2 Caracterização dos padrões de mobilidade	99
<i>6.2.1 Análise descritiva dos padrões</i>	99
<i>6.2.2 Análise descritiva dos padrões</i>	102
<i>6.2.3 Análise de evolução temporal</i>	111
<i>6.2.4 Análise de evolução espacial</i>	113
7 CONCLUSÕES E RECOMENDAÇÕES	119
7.1 Conclusões	119
7.2 Limitações e recomendações	121
REFERÊNCIAS	124
APÊNDICE	133

1 INTRODUÇÃO

1.1 Problemática

A cidade Fortaleza vem sendo reconhecida, nos últimos anos, por suas políticas de planejamento urbano, especificamente, nas áreas de mobilidade. Diversos investimentos e melhorias na oferta no sistema de transporte público ocorreram com a intenção de melhorar a experiência dos usuários desses sistemas. Um exemplo prático desse reconhecimento, foi a premiação 2019 Sustainable Transport Award, prêmio de uma competição global de projetos e intervenções inovadoras que promovem a mobilidade urbana de forma sustentável e inclusiva (ITDP, 2018), garantida à cidade de Iracema em 2018.

Sobre essas mudanças, Sousa (2019) traz algumas intervenções que ocorreram entre os anos de 2014 a 2019 no sistema de transporte público de ônibus em Fortaleza as quais merecem destaque no processo de consolidação de um sistema público de melhor qualidade e, possivelmente, com melhor níveis de acesso, mas não necessariamente menos desigual: a) implantação de linhas que operam com veículos biarticulados em um sistema BRT; b) ampliação da quantidade de faixas exclusivas de ônibus, capazes de reduzir as velocidades médias de deslocamentos nos principais corredores arteriais da cidade; e c) implantação de sistema de bicicleta compartilhado, que se apresenta como mais uma alternativa para o acesso e difusão do transporte público.

Além disso, os usuários do sistema contam, desde 2013, com a política de Bilhete Único, que permite a integração entre viagens fora dos terminais dentro de um período de duas horas. Apesar do investimento na melhoria da oferta, os dados analisados nesta pesquisa apresentaram uma tendência de **queda na demanda do sistema de transporte público por ônibus em Fortaleza nos últimos anos.**

Entende-se que a queda dessa demanda pode ter acontecido por diversas razões. Balcombe *et al.* (2004), por exemplo, cita diversos fatores que impactam na demanda por Transporte Público: Tarifa; Renda e Posse de automóveis; Tempo de viagem; Microacessibilidade; Tempo de caminhada; Tempo de espera; Tempo de integração; Confiabilidade do sistema (horários); Disponibilidade de modos alternativos; Fatores socioeconômicos; Uso do solo;

Políticas públicas; Segurança; Qualidade do veículo; Conveniência do modo ao usuário. Esse conjunto de fatores torna o processo de análise e de planejamento da demanda bastante complexo.

A queda na demanda do sistema de transporte público por ônibus em Fortaleza é uma problemática citada por Henrique (2004) e Sousa (2019), que atribuem esta queda aos baixos índices de acessibilidade da população de baixa renda e ao aumento da posse de veículo por parte da população. Porém, como veremos no decorrer desta pesquisa, a não consideração da heterogeneidade da demanda e das suas variações espaço-temporais na revisão de literatura são lacunas relevantes que indicam que este problema ainda pode ser investigado a partir de outros métodos.

No modelo conceitual do planejamento de transportes de Soares (2014), a investigação desses problemas pertence a uma etapa chamada caracterização. A caracterização do fenômeno da demanda em sistemas de transportes apresenta uma complexidade de análise inerente ao sistema analisado, uma vez que milhares de viagens acontecem diariamente (MEYER; MILLER; 2001).

Por exemplo, em Fortaleza, o sistema de ônibus possui em média de um milhão de viagens realizadas por dia. Em um ano de análise, isso configura uma grandeza de, em média, trezentos milhões de viagens das quais são realizadas por diferentes usuários, levando essa demanda a ter diferentes características espaciais e temporais. Além disso, cada usuário do sistema possui diferenças em suas características socioeconômicas, o que leva a cada viagem possuir características únicas. O conhecimento sobre as variações temporais e espaciais, assim como as características socioeconômicas atreladas às viagens são essenciais para o planejamento adequado de um sistema de transporte (ORTÚZAR, 2011).

De acordo com Soares (2014), a etapa de caracterização é dependente da disponibilidade de dados que sejam capazes de fornecer informações suficientes ao propósito do processo. Nesse sentido, pesquisas domiciliares e pesquisas de origem-destino, por fornecerem uma quantidade razoável de informações sobre viagens, atividades realizadas e dados demográficos da população da amostra pesquisada, são bastante utilizadas no processo de planejamento de transportes, principalmente nas etapas de caracterização e diagnóstico.

Porém, essas pesquisas apresentam algumas limitações para serem realizadas, como a demanda de um alto custo financeiro e de tempo de coleta e processamento dos dados, inviabilizando a realização delas de forma frequente. Além disso, essas pesquisas, normalmente,

coletam informações de um ou dois dias de viagem e somente são atualizadas depois de muitos anos (ANDA; FOURIE; ERATH, 2017). Criando assim, retratos estáticos das viagens realizadas. Devillaine, Munizaga e Trépanier (2012) defendem que a utilização de pesquisas domiciliares para o planejamento da demanda introduz vieses e erros, principalmente devido a alterações subsequentes em qualquer sistema que afeta a percepção da população, podendo levar a uma mudança nas variáveis pesquisadas e nas preferências de cada usuário, portanto, podem diminuir a validade dos dados da pesquisa. Essa característica torna as informações coletadas confiáveis apenas enquanto as mudanças não ocorrerem.

Nesse sentido, os retratos representados pelas pesquisas domiciliares apresentam dificuldade em perceber e transmitir a variabilidade e as características de heterogeneidade dos padrões de viagens. Além disso, normalmente, os estudos que utilizam dados de pesquisas domiciliares tendem a caracterizar essa demanda ou a mobilidade dos usuários, através de indicadores agregados por zonas, assumindo a premissa que toda a população de uma zona pode ser representada pelo seu valor médio. Essa premissa vai contra o fenômeno de heterogeneidade da demanda. Portanto, alguns autores defendem que as análises e os planejamentos de transportem devem considerar essa heterogeneidade. Hawkins e Habib (2020) sugerem que seja evitado a atribuição de um único valor de mobilidade à população heterogênea de uma grande região. Wang, Wang e Choudhury (2019) aconselham que os planos precisam ser capazes de atender às diferentes preferências de todos os grupos de usuários.

Como forma de contornar as limitações das pesquisas domiciliares e, conseqüentemente, dos estudos que caracterizam a demanda de transporte a partir destas pesquisas, novas fontes de dados vêm sendo utilizadas, esses dados são oriundos de sistemas de tecnologia de informação. Os sistemas automáticos de bilhetagem (*Smart Card*), de localização geográfica veículos (GPS) e Especificação Geral de Feed de Trânsito (GTFS) são tecnologias que se destacam por gerarem uma base de dados de larga escala quase que populacional da demanda e oferta de um sistema de transporte público, sendo assim, nos últimos anos, bastante utilizadas no planejamento de transportes (PELLETIER; TRÉPANIER; MORENCY, 2011).

Esses novos dados apresentam várias vantagens sobre pesquisas tradicionais, como a coleta contínua e o aumento significativo no tamanho da amostra dos dados de *Smart Card* que permitem observar e interpretar padrões ao longo do tempo, ao contrário das informações estáticas de pesquisas convencionais (DEVILLAINÉ; MUNIZAGA; TRÉPANIER, 2012).

A partir das observações, praticamente, populacionais dos dados de *Smart Card* permitem o cálculo de indicadores, possibilitando a consideração da heterogeneidade da demanda para etapa de caracterização. Dessa forma, a investigação de características da demanda a partir desses dados aparenta ser uma alternativa perante às limitações dos dados de pesquisas domiciliares.

Porém, mas algumas dificuldades ainda estão presentes nesses dados. Uma delas é a enorme quantidade de usuários e conseqüentemente a enorme quantidade de viagens realizadas nesse sistema, tornando inviáveis as análises de demanda a nível de agregação individual. Como alternativa, alguns autores utilizam técnicas de clusterização para identificar grupos de usuários que possuam características de mobilidade e socioeconômicas em comum (AGARD; PARTOVINIA; TRÉPANIER, 2013; PIRERONI, 2018; ZHAO *et al.*, 2017). Esses autores defendem que o agrupamento de usuários a partir de características em comum tem o potencial de ajudar os planejadores de transporte público a entenderem o comportamento de seus clientes, podendo oferecer um serviço mais adequado de acordo com a demanda.

A identificação destes grupos a partir dessas características implica dizer que, assim como a demanda do sistema, estes grupos também podem estar distribuídos de forma diferente no espaço e no tempo. Essa segregação da demanda em grupos pode ajudar no processo de compreensão do porquê a demanda está caindo ao longo do tempo, uma vez que a demanda possui um comportamento heterogêneo e que esta queda pode ser mais intensa em um grupo do que em outro. Além disso, este método também permite que a investigação da dinâmica desses grupos ao longo do tempo, uma vez que os dados de *Smart Card* possuem observações contínuas.

Assim, nossa investigação observa as restrições existentes dos tipos de dados obtidos pelas pesquisas tradicionais e a disponibilidade de dados de novas tecnologias. Sem abandonar as dificuldades na avaliação da demanda em nível individual e a necessidade de compreender melhor as variações na demanda do sistema de transporte público por ônibus em Fortaleza com foco no problema de queda na demanda. A partir disso, nosso problema de pesquisa reside em, através dos dados de *Smart Card*, investigar a existência de grupos de usuários com comportamentos homogêneos de utilização do sistema, como esses grupos são distribuídos espacialmente e temporalmente e como a queda na demanda impacta esses grupos.

1.2 Objetivos

Dada a problemática apresentada, o objetivo geral deste estudo é: **Caracterizar os padrões de mobilidade dos usuários do sistema de transporte público por ônibus utilizando *Big Data*, relacionando estes padrões com os problemas de mobilidade.** Para isso, serão aplicadas técnicas de análise descritiva simples e espacial, a fim de compreender como os padrões identificados na demanda variam no espaço e no tempo, assim como, quais são os perfis de usuários que mais estão sofrendo variações durante o período analisado. Para que esse delineamento de pesquisa seja alcançado, instituímos os seguintes objetivos específicos:

- identificar e interpretar os padrões de mobilidade dos usuários do sistema de transporte público de ônibus;
- caracterizar os padrões identificados no espaço e sua evolução temporal para um período de cinco anos;
- avaliar como os padrões de mobilidade identificados se relacionam com os problemas de mobilidade.

1.3 Estrutura da dissertação

Apresentamos, nesta dissertação, 7 capítulos em que desenvolvemos a temática, os quais estão dispostos a seguir: O capítulo 1 (Introdução) apresenta as lacunas, os problemas de pesquisa, os objetivos e estrutura da pesquisa.

Já Capítulo 2 (Caracterização da Mobilidade) discute as diferenças conceituais e práticas entre mobilidade e acessibilidade, assim como a relação desses dois temas com o planejamento de transporte público. Também é levantado um histórico de estudos de caracterização da mobilidade realizados em Fortaleza. Além de ser definido o significado de padrão de mobilidade para esse estudo.

No Capítulo 3 (Caracterização e identificação de padrões de mobilidade a partir do *Big Data* de Transporte Público) apresentamos as novas tecnologias responsáveis pela a geração de imensos bancos de dados e em como essas novas fontes de dados podem ser úteis para o planejamento de transportes. Também realizamos um levantamento histórico dos estudos recentes que utilizaram dados de *Big Data* do Transporte Público de Fortaleza. Esse capítulo traz ainda,

uma rápida discussão sobre o que é um padrão de mobilidade e em como os padrões podem ajudar no processo de caracterização da demanda. É apresentado também as ferramentas mais utilizadas na literatura para identificação de padrões de mobilidade através de dados de *Smart Card*, assim quais as melhores práticas identificadas de aplicação desses métodos.

Nossa discussão, no Capítulo 4 (Contextualização dos Sistemas Analisados), se propõe a angariar um conjunto de informações relevantes para a construção das hipóteses de pesquisas desta pesquisa. Essas informações são resultado de estudos que se propõem a caracterizar aspectos essenciais dos Sistemas de Transporte, Uso do Solo e de Atividades de Fortaleza.

O capítulo 5 (Proposta Metodológica de Caracterização de Padrões de Mobilidade) propõe um método de identificação de padrões de mobilidade através de uma ferramenta de clusterização (*k-means*) que utiliza atributos de mobilidade para classificar os usuários em diferentes padrões. Em seguida, é proposto um método de caracterização dos padrões de mobilidade e da evolução desses padrões entre os anos de 2014 e 2018.

Os resultados da aplicação da metodologia proposta serão discutidos no capítulo 6 (Resultados). Este capítulo traz análises e discussões do processo de identificação dos padrões e de caracterização dos padrões e da evolução entre os anos de 2014 e 2018.

O capítulo 7 (Conclusões e Recomendações) aborda as conclusões sobre o método aplicado e como este contribui para preencher as lacunas identificadas na revisão de literatura e em verificar as hipóteses levantadas. Será também proposto algumas recomendações para trabalhos futuros, com base, principalmente, nas limitações identificadas deste estudo.

2 CARACTERIZAÇÃO DA MOBILIDADE

Neste capítulo será discutido sobre as definições de mobilidade e acessibilidade, bem como esses temas são abordados no planejamento de transportes público. Além disso, será discutido como a mobilidade vem sendo caracterizada em Fortaleza e quais as lacunas identificadas.

2.1 Conceitos e práticas (mobilidade x acessibilidade)

Mobilidade e acessibilidade são dois tópicos amplamente discutidos no planejamento de transportes e, muitas vezes, suas concepções se cruzam e acabam sendo confundidas, sendo essencial apontar suas diferenças. Assim, Garcia (2016) faz um levantamento das definições dos dois tópicos (Figura 1) e resume que, no geral, as definições de mobilidade se referem a habilidade de movimentação do indivíduo, enquanto as definições de acessibilidade referem-se à facilidade de acesso ou de interação com o destino. As interpretações dessas definições de literatura não são claras e Garcia (2016) também traz essa discussão.

Em sua pesquisa, Garcia (2016), a partir de uma revisão realizada, define que a mobilidade expressa a facilidade de movimento e está relacionada com a capacidade individual de se deslocar de um lugar para outro, enquanto a acessibilidade representa a facilidade de chegar ao destino / atividades e, portanto, de interagir, sendo influenciada pelas características dos sistemas de transportes e de uso do solo. Essa definição avança ao deixar clara a relação da acessibilidade, não somente com o sistema de transportes, mas também com sistema de uso do solo. Será essa definição que utilizaremos neste trabalho.

Alguns autores citam ainda, diferentes tipos de acessibilidade. Assim, se faz necessário para estudos que se propõem a analisar a demanda de um sistema de transporte público, que algumas dessas diferenças sejam destacadas, por exemplo, Cardoso (2008) que discute dois tipos de acessibilidades: a acessibilidade ao sistema de transporte (mede a facilidade de o usuário acessar o sistema de transporte coletivo em sua região de moradia, trabalho etc.) e a acessibilidade a destinos (mede, após o acesso ao sistema de transporte, a facilidade de se chegar ao destino desejado).

Figura 1 – Definições de Acessibilidade e Mobilidade

Accessibility	Mobility
<ul style="list-style-type: none"> • Potencial de oportunidades de interação (Hansen, 1959) • Potencial de mobilidade (Jones, 1981) • Facilidade para realizar viagens de / para certas zonas de tráfego (Ortúzar e Willumsen, 1994) • Facilidade de acesso de um ponto a outros pontos no rede (Taaffe et al, 1996) • Capacidade de viajar entre diferentes atividades (Vuchic, 2000) • Facilidade para chegar ao destino / atividades (Levine e Garb, 2002) • Até que ponto o sistema de uso da terra e transporte permite que os indivíduos alcancem as atividades por meio de meios de transporte (Geurs e van Wee, 2004) • Propriedade do meio de transporte (Magalhães et al., 2013) 	<ul style="list-style-type: none"> • Potencial de movimento (Hansen, 1959) • Capacidade dos indivíduos de se deslocarem de um lugar para outro influenciada pelo desempenho do transporte, bem como por características socioeconômicas (Tagore e Skidar, 1995) • Facilidade de movimento (Levine e Garb, 2002) • Facilidade de movimento ou habilidade de mover-se ou ser movido (Handy, 2005) • Capacidade de indivíduos ou grupos de se moverem fisicamente de um lugar para outro (EEA, 2006) • Propriedade de algo que pode ser transportado (Magalhães et al., 2013)

Fonte: Adaptado de Garcia (2016).

De modo oposto, aos conceitos e suas interpretações, a prática de planejamento da mobilidade e acessibilidade possuem pilares diferentes. Sousa (2019) explica essas diferenças e descreve a existência de três paradigmas no planejamento de transportes, que são originalmente expostos por Banister (2008).

O primeiro paradigma, muito utilizado décadas atrás, focava em problemas operacionais de transportes (congestionamento, atrasos, acidentes etc.) e se tratava de uma interpretação funcional do sistema, baseado em tráfego (LOPES; CAVALCANTE; LOUREIRO, 2018). Esse paradigma levou ao aumento de investimentos em infraestrutura (oferta) como solução principal dos problemas de transportes, trazendo consequências para o comportamento dos usuários que podem ser vistas até hoje, como o crescente aumento na dependência de veículos motorizados individuais (MEYER; MILLER, 2001).

O segundo paradigma descrito por Sousa (2019) se trata do planejamento da mobilidade, em que se percebeu-se a necessidade de enfatizar os indivíduos que utilizavam o sistema. Sousa (2019) aponta as viagens como maior preocupação desse planejamento, sem dar muita importância aos motivos que geram essas viagens, bem como, demonstra como o planejamento se baseia no desequilíbrio entre oferta e demanda, propondo soluções que garantam mais quantidade de viagens.

Em relação a esse paradigma, Banister (2008) aponta a existência de dois princípios fundamentais: o primeiro princípio diz que a viagem é uma demanda derivada e não uma atividade

que as pessoas desejam realizar por si mesma. Ou seja, é o valor da atividade destino que gera viagem, já o segundo princípio aponta que as pessoas minimizam seus custos generalizados de viagem, operacionalizados, principalmente, por meio de uma combinação dos custos da viagem e do tempo gasto para viajar.

Esses dois princípios básicos têm importantes consequências, uma vez que estão incorporados na maioria dos estudos de análise e avaliação dos planejamentos de transportes. Eles ajudam a explicar a predominância de soluções de transporte para problemas urbanos e o enorme crescimento em viagens mais rápidas e de longas distâncias, já que o aumento da velocidade das viagens compensou o aumento dos custos das viagens. Por fim, Banister (2007) defende também que esses princípios levam o transporte público local, bicicletas e caminhadas para uma posição menos atraente para os usuários, e que, conseqüentemente, resulta também no maior uso do automóvel individual.

O Terceiro e o mais recente paradigma, é o planejamento da acessibilidade (GARCIA, 2016; SOUSA, 2019) ou planejamento da mobilidade sustentável (BANISTER, 2008). Atualmente, o cenário de planejamento de transportes foca grande parte dos seus esforços nesse paradigma. Sousa (2019) descreve-o como um paradigma que apresenta diferenças conceituais, morais e metodológicas, resultando em uma proposta de abordagem de planejamento que destaca a importância e a combinação de novos valores e princípios como acessibilidade, equidade e sustentabilidade (BANISTER, 2008; HANDY, 2016; KENWORTHY, 2007; MACÁRIO, 2016).

Esse paradigma tem como propósito promover socioespacialmente a região planejada de modo a contribuir para a qualidade de vida e justiça social da população. Para que o desenvolvimento socioespacial da região planejada ocorra, é importante que os subsistemas urbanos, que influenciam nos problemas analisados, sejam reconhecidos e planejados de forma integrada (LOPES, 2015; GARCIA, 2016; SOUSA, 2019).

2.2 Problemas de mobilidade

O princípio da sustentabilidade, uma das bases do planejamento da acessibilidade, é normalmente descrito como uma estrutura que possui três dimensões: econômica, social e ambiental (BANISTER, 2008). A partir dos aspectos sociais, a equidade é vista como mais um de seus componentes e é um dos que têm recebido menos atenção tanto na teoria quanto na prática

do planejamento, quando comparada às outras dimensões (GARCIA *et al.*, 2018; BOSCHMANN; KWAN, 2008; MARTENS, 2006; UTENG, 2007; VALLANCE *et al.*, 2011). Entretanto, há autores defendendo que para se alcançar a sustentabilidade deve-se existir um nível mínimo social, sendo a equidade o princípio substancial por trás da sustentabilidade (GARCIA *et al.*, 2018; BEDER, 2000; FEITELSON, 2002; WEISS, 1992).

Garcia *et al.* (2018) propõem, então, a utilização das teorias éticas do suficientarismo e igualitarismo para construir uma estrutura de definição e análise de problemas de acessibilidade e mobilidade. A teoria do suficientarismo, segundo Meyer e Roser (2009), assume que todos devem estar “bem de vida” e defende o estabelecimento de um limiar de bem-estar abaixo do qual os indivíduos podem ser considerados insatisfeitos ou mal atendidos. Já o igualitarismo considera que todos devem ser tratados de forma igual e, portanto, está relacionado com as diferenças e não com os valores absolutos de bem-estar, portanto, defende que as diferenças relativas entre os estados das pessoas precisam ser eliminadas ou reduzidas.

Garcia *et al.* (2018), a partir dessas duas teorias, propõem uma outra abordagem, em que as diferenças devem ser eliminadas ou reduzidas considerando uma abordagem equitativa e não igualitária, ou seja, maiores benefícios devem ser dados aos mais desfavorecidos, a fim de alcançar um estado justo para todos.

Garcia *et al.* (2018), com base nas considerações feitas a partir das teorias éticas do suficientarismo e igualitarismo, propõem uma tipologia de problemas de acessibilidade e mobilidade que conta com quatro categorias principais (Figura 2): distribuições desiguais, injustas, inadequadas e insustentáveis de acessibilidade e mobilidade urbana. A tipologia de problemas proposta por Garcia *et al.* (2018) é importante, pois ajuda a definir e esclarecer quais os tipos de problemas de mobilidade e acessibilidade que serão investigados nessa pesquisa.

Figura 2 – Classificação dos problemas de acessibilidade e mobilidade



Fonte: Adaptado de Garcia (2016) e Garcia *et al.* (2018).

Conforme exposto, originalmente, a acessibilidade tem um conceito intimamente relacionado à mobilidade e vem sendo usada no planejamento e na avaliação do sistema de transporte por meio de abordagens baseadas nas atividades. Os dois fatores estão diretamente ligados um ao outro, logo, as políticas que promovem a mobilidade geralmente também promoverão a acessibilidade e farão com que os destinos sejam mais fáceis de alcançar. Porém, é possível ter boa acessibilidade com mobilidade deficiente e vice-versa (HANDY, 2002).

Para compreender melhor essa relação, é preciso entender a mobilidade como resultado da escolha do indivíduo de realizar uma viagem. Já a acessibilidade precisa ser entendida como resultado da relação entre a oferta dos sistemas de transportes acessíveis ao usuário com a distribuição das residências e das atividades no uso do solo.

Sendo assim, por mais que o usuário tenha acesso às atividades a partir dos sistemas de transportes ofertados, se o indivíduo não precisar ou decidir por não realizar viagens, o seu nível de mobilidade será baixo mesmo possuindo níveis considerados de acessibilidade. Esse exemplo ajuda a entender quais problemas de mobilidade e acessibilidade são diferentes e quais, ocasionalmente, precisam ser investigados de forma separada.

Conforme apresentado na introdução, o problema principal e motivacional que essa pesquisa se propõe a investigar é a queda na demanda de transporte público por ônibus em

Fortaleza. A mobilidade é, muitas vezes, mensurada na literatura por indicadores orientados pela demanda (GARCIA, 2016). Portanto, problemas de queda na demanda são naturalmente associados a problemas de mobilidade, essa associação permite apontar quais os tipos de problemas que, a partir da tipificação de Garcia *et al.* (2018), serão investigados nesta pesquisa: Distribuição desigual, e insustentável da mobilidade.

2.3 Características que impactam na acessibilidade e na mobilidade

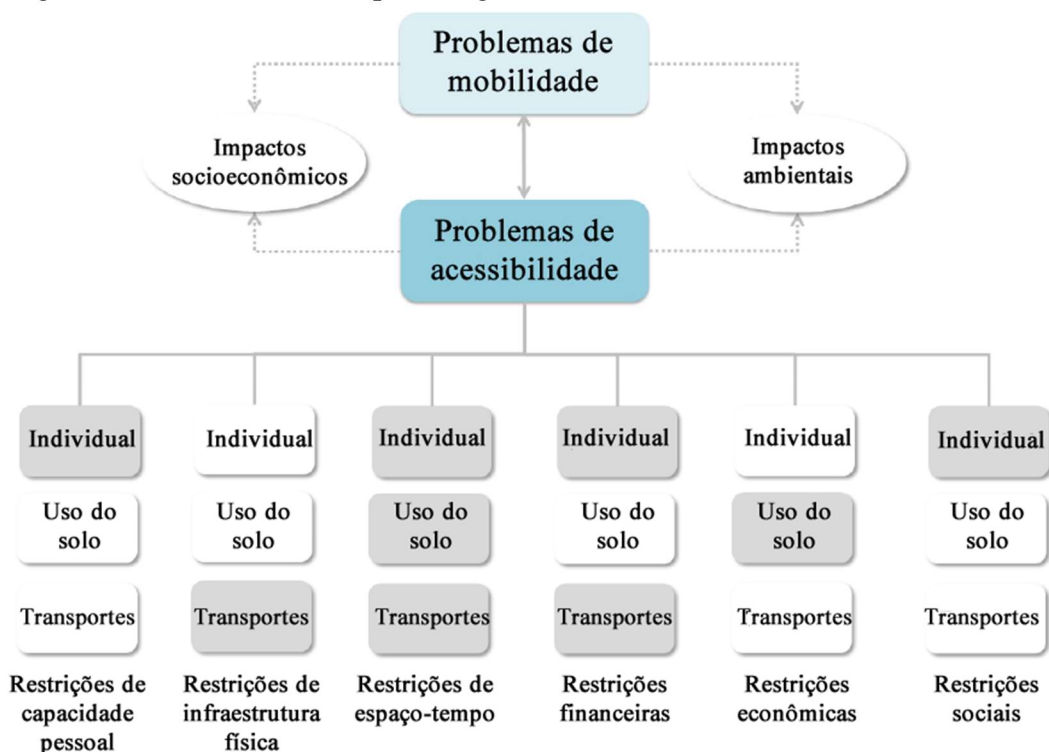
Um dos fatores importantes na acessibilidade e na mobilidade é o sistema de transporte, que é descrito por Lopes (2015) como uma rede de transporte que surge a partir de uma necessidade de locomoção de pessoas e cargas. Essa rede tem uma capacidade máxima de utilização, limitando-a a partir de características físicas (infraestrutura) e operacionais (lógicas) do sistema. A necessidade de transportes leva os atores a realizarem decisões de viagem que envolvem escolhas de modo de transporte ou itinerários.

Nesse sentido, Lopes (2015) defende que existe pelo menos três elementos distintos que afetam as decisões de viagem, sendo eles: a distribuição espacial do uso do solo, a distribuição das oportunidades (trabalho, educação e outras atividades) e o desempenho do sistema de transportes ofertados. O resultado dessas escolhas forma as linhas de desejo, que são, basicamente, a materialização da quantificação dos fluxos desejados na rede, associando origens, destinos, modos de transporte e as rotas. A não efetivação dessas linhas de desejos por parte dos indivíduos pode ocorrer por diversos tipos de restrições (MENEZES, 2015; GARCIA; 2016). Essas restrições e suas relações causais com os problemas de acessibilidade e mobilidade podem ser visualizadas na Figura 3.

A partir de Menezes (2015), entende-se que as restrições de acessibilidade podem resultar tanto na diminuição da mobilidade, quanto na diminuição dos níveis de atividades realizados pelo indivíduo. Portanto, essas restrições devem ser compreendidas como características importantes que determinam como a mobilidade de um indivíduo é moldada. De acordo com Garcia (2016), essas restrições podem ser resumidas e caracterizada em três grupos: inadequações na distribuição e no uso do solo, características individuais e ineficiências na oferta do sistema de transporte. Meurs e Van Wee (2003) também apresentam as características do usuário, os aspectos do sistema de transporte e do sistema de uso do solo como as principais

influências nos padrões de viagens. Essas três categorias são importantes para compreender que, independentemente da restrição analisada, microscopicamente, a mobilidade dos usuários é impactada por cada restrição de maneira diferente e individual.

Figura 3 – Relações causais por categorias de restrições de acessibilidade



Fonte: Adaptado de Garcia (2016) e Menezes (2015).

Por exemplo, um indivíduo é impactado pelas restrições de inadequações na distribuição e no uso do solo quando necessita realizar uma atividade que não está acessível ou que haja certa dificuldade de acesso. Nesse mesmo exemplo, um outro indivíduo que resida próximo ou na mesma residência pode não sofrer por esse tipo de restrição se este necessitar realizar uma atividade diferente que seja mais acessível. Em relação às restrições dadas pelas características individuais, esse impacto é ainda mais singular, por mais que as decisões dos usuários sejam moldadas pela infraestrutura ofertada e pelo uso do solo, as necessidades de cada um são diferentes, essas necessidades acarretam que suas decisões de viagem, e consequentemente sua mobilidade, também sejam singulares.

Da mesma maneira acontece com as restrições relacionadas a ineficiências na oferta do sistema de transporte, de forma microscópica, cada usuário é afetado diferentemente pelas ineficiências na oferta de um sistema de transporte, já que as necessidades de realizar atividades e as decisões de viagem são diferentes para cada indivíduo. Essa discussão é importante para este trabalho, pois ajuda a entender que indivíduos distintos possuem padrões de mobilidade diferentes, e assim, questionar o método de caracterização da mobilidade em que usuários de uma mesma zona são representados por um mesmo nível de mobilidade.

Para cada uma dessas restrições, várias características podem ser identificadas, como: posse de veículos, endereço de residência, renda, gênero, etnia, idade, emprego e nível educacional são características que fazem parte das restrições das individuais. Já as características das ineficiências na oferta de transporte são: oferta de diferentes modos, tempo de espera, lotação, eficiência e abrangência das rotas, oferta de paradas no território, tarifa e nível de conforto dos veículos. No tocante às inadequações na distribuição e no uso do solo o mix de uso do solo e a distância percorrida para acesso as atividades são as principais características.

Alguns autores, baseados em teorias de psicologia social, defendem ainda que o comportamento de viagem é moldado por um componente racional, bem como um componente irracional. Acker, Van Wee e Witlox (2010) propõem um modelo conceitual onde, inicialmente, é considerado o comportamento de viagem como derivado de decisões de atividades de curto prazo, decisões de localização de médio prazo e decisões de estilo de vida de longo prazo. Em seguida, decisões comportamentais de viagens são consideradas como o resultado de uma avaliação entre influências racionais e irracionais. Desta forma, o modelo conceitual concentra-se no comportamento de viagem do indivíduo. Porém, os autores entendem que o indivíduo pertence a uma rede social de família, amigos e colegas e mora em uma determinada vizinhança que pode afetar o comportamento do indivíduo, assim, esses conceitos também são incluídos no modelo. Por fim, o modelo como um todo é colocado dentro de (1) um nível individual, (2) um ambiente social e (3) um ambiente espacial.

Para Acker, Van Wee e Witlox (2010), embora algumas das decisões de viagens sejam motivadas ou impactadas pelo estilo de vida do indivíduo, usuários socioeconomicamente homogêneos podem realizar decisões comportamentais diferentes. Isso pode ser devido às percepções, atitudes e preferências individuais em relação ao local, atividade e comportamento de viagem (VAN WEE, 2002; MOKHTARIAN; CAO, 2008). A existência de decisões

comportamentais diferentes em um mesmo grupo social reforça a existência de uma demanda heterogênea dentro de uma mesma região ou dentro de mesmo grupo social.

2.4 Indicadores de mobilidade

Indicadores de mobilidade é assunto de diversos estudos e pesquisas (COSTA, 2008; COSTA; NETO; BERTOLDE, 2017; GILLIS; SEMANJSKI; LAUWERS, 2015). A utilização de indicadores capazes de representar os problemas relacionados à mobilidade e acessibilidade é essencial no processo de planejamento, principalmente, na etapa de caracterização (SOARES, 2014).

Por isso, Magalhães (2004) descreve os indicadores como elementos centrais nos processos de decisão, esses indicadores permitem sintetizar e representar um fenômeno que, muitas vezes, não é possível através apenas de dados brutos. É a partir dos indicadores que os decisores tomam conhecimento de informações essenciais ao processo de planejamento e, a partir deles, constroem estratégias para analisarem desempenho, eficiência, eficácia, efetividade e alcançar os objetivos. Litman (2007) apresenta uma revisão de literatura em que se destaca alguns indicadores de mobilidade e os classifica em três tipos: Econômico (viagens diárias per capita, divisão modal, tempo médio de deslocamento, velocidade média da viagem, custos de congestionamento, custos totais por viagem), Social (acidentes e fatalidades per capita, qualidade do transporte para pessoas mais necessitadas, orçamento dedicado ao transporte por família, Índice de satisfação geral com o sistema de transporte, micro acessibilidade) e Ambiental (consumo de energia per capita, consumo de energia por tonelada de carga, emissões de poluentes de ar per capita desagregadas por modo, uso da terra dedicado a meios de transporte, exposição à poluição atmosférica e sonora).

Saloman e Moktharian (1998) classificam os indicadores de mobilidade entre duas perspectivas: demanda e oferta. Os indicadores de oferta são subdivididos em indicadores de Facilidade de Viagem e Potencial de Viagem. Os indicadores do tipo facilidade de viagem são indicadores de mobilidade que retratam as alternativas de viagem disponíveis (propriedade de carro, disponibilidade de modos alternativos como sistemas de compartilhamento de bicicletas e carros etc.) ou o desempenho do sistema de transporte (velocidade média por modal, relação volume / capacidade ou veículo-km).

Os indicadores de potencial de viagem são indicadores que retratam a capacidade individual de viajar, através desse tipo de indicador é possível calcular o potencial de mobilidade dos indivíduos a partir de indicadores que mensurem a oferta de sistemas de transportes em determinado espaço-tempo e o tempo disponível para realizar uma viagem. Garcia (2016) realiza uma discussão que esse tipo de indicador, apesar de categorizado por Saloman e Moktharian (1998) como indicadores de mobilidade, são indicadores que representam mais a acessibilidade do que a mobilidade.

Já os indicadores de mobilidade orientados pela demanda representam o comportamento de viagem dos indivíduos, esses indicadores procuram mensurar a quantidade ou duração das viagens realizadas. Para facilitar a compreensão e apresentação desses indicadores, eles costumam ser classificados a partir das características da viagem (modo, finalidade, duração, distância percorrida, origem, destino, horário realizada) e socioeconômicas dos indivíduos (renda, sexo, idade).

Vale ressaltar que a avaliação de potenciais problemas através de um único indicador não é indicada pela literatura. Isso ocorre quando, por exemplo, a quantidade de viagens é alta e o nível de mobilidade pode ser considerado bom, aparentemente, não é percebido nenhum problema. No entanto, se a duração ou distância dessas viagens também for alta, talvez um problema de baixo nível de acessibilidade possa existir e não ser percebido. Sendo necessário, então, a combinação de diferentes indicadores que possam refletir melhor os problemas analisados (SALOMAN; MOKTHARIAN, 1998; GARCIA, 2016).

Os três tipos de indicadores apresentados são importantes para a compreensão dos fenômenos de mobilidade em um sistema de transportes, porém, os indicadores de demanda apresentam uma maior facilidade de entendimento, pois, normalmente, representam viagens que já foram realizadas e são, facilmente, comparáveis para diferentes grupos e características de indivíduos.

No entanto, Garcia (2016) defende que indicadores orientados pela oferta, embora não possibilitem uma imagem precisa do comportamento de mobilidade em si, podem ser úteis para entender as razões que levam a certos padrões de mobilidade, por exemplo, a propriedade de veículos pode levar a mais mobilidade de carros particulares enquanto número ou frequência de linhas de metrô pode explicar o porquê certos contextos urbanos apresentam uma mobilidade de

transporte público mais intensa do que outros. Neste trabalho, serão utilizados indicadores orientados pela demanda.

2.5 O papel do transporte público no planejamento da mobilidade

Políticas que promovem o transporte público são essenciais para garantir uma sociedade mais justa e equânime, uma vez que estes assumem papel central no processo de garantir a população mais vulnerável, maiores níveis de mobilidade e melhores níveis de acessibilidade, o que na prática significa garantir a essa parcela da população mais acesso a oportunidades de trabalho, como também a serviços de saúde, educação e lazer. Para Hickman, Hall e Banister (2013), as principais políticas de transporte para o planejamento urbano, que estão em linha com os paradigmas da mobilidade sustentável, para os próximos anos devem ser: a) Investimento em transporte público; b) Redução da emissão de CO₂ dos veículos automotivos; c) Investimento em infraestrutura para caminhada e ciclismo.

Em estudo recente, o Instituto de Pesquisa Econômica Aplicada (IPEA) apontou que em cidades de pequeno, médio e grande porte no Brasil, em média, 65,9% da população utiliza o transporte público com modo principal para realizar suas atividades (IPEA, 2019). Os dados apontados pelo IPEA dão sinais dos impactos e da importância da realização de políticas públicas que visam promover o transporte público e promover a mobilidade e acessibilidade da população, ou seja, essas políticas têm grande potencial em atuar sobre os problemas tipificados por Garcia *et al.* (2018).

Dada a relevância do transporte público para a sociedade, surge também a importância de planejar esse sistema a partir dos fundamentos do paradigma do planejamento da acessibilidade previamente apresentado. De acordo com Mugion *et al.* (2018), usuários que possuem o seu próprio automóvel estão menos propícios a utilizar o transporte público do que os usuários que não possuem, considerando esse fato, bem como o desuso dos veículos particulares voluntariamente a partir da conscientização particular do indivíduo (TERTOOLEN *et al.*, 1998), para reduzir a dependência do carro, é necessário promover medidas eficazes, como a melhoria nos níveis de qualidade do serviço, satisfação geral e consciência ambiental.

Diante disso, a literatura da área mostra que o cenário ideal para a sustentabilidade do meio urbano é o crescimento da infraestrutura e da utilização do transporte público. Porém, assim

como em Fortaleza, o cenário nas cidades brasileiras é de queda expressiva na quantidade de usuários que utilizam o transporte coletivo, principalmente em grandes centros urbanos. Segundo a Associação Nacional das Empresas de Transportes Urbanos (NTU, 2018), em 5 grandes capitais brasileiras, entre 1997 e 2017, houve redução acumulada de aproximadamente 35,6% de usuários, sendo essa queda mais acentuada nos últimos 5 anos.

Muitas empresas operadoras e órgãos públicos responsáveis por gerir os sistemas de transportes coletivos ainda estão presas nos antigos paradigmas de planejamento de transportes em que o foco do planejamento é a oferta do sistema (IPEA, 2003). Portanto, para buscar a implementação de soluções no planejamento da acessibilidade é necessário um envolvimento maior das partes interessadas, buscando oferecer um serviço de melhor qualidade que realmente atenda às necessidades da população e que se enquadre nas políticas do Plano Nacional de Mobilidade Urbana, Lei Federal n. 12.587, de 2012.

Apesar do transporte público já ser por natureza uma política inclusiva, o debate e as ações sobre como aprimorar o seu planejamento com objetivo de garantir um serviço de transporte mais equânime ainda é insuficiente dado o seu potencial. Nesse sentido, o planejamento dos sistemas de transportes públicos, a partir da identificação dos problemas tipificados por Garcia *et al.* (2018), pode ser um caminho a ser percorrido para alcançar esses objetivos.

Algumas perguntas podem ser feitas para verificar se as políticas de transporte público estão atuando de forma a minimizar os problemas de mobilidade: A oferta dos sistemas de transporte público vem sendo planejadas para melhorar os níveis de mobilidade da população de baixa renda e que vive em zonas periféricas e com menos acesso? A sustentabilidade, ou seja, os níveis de mobilidade e acessibilidade vem melhorando ao longo do tempo para os usuários de transporte público? As reais necessidades da população são identificadas a fim de ofertar um sistema de transporte que atenda essas necessidades?

2.6 Caracterização da mobilidade em Fortaleza

A caracterização é uma das etapas do processo de planejamento de transportes proposto por Soares (2014) que pode ajudar a responder as questões levantadas no tópico anterior. A caracterização tem como finalidade a estruturação das percepções dos atores com dados que sejam capazes de expressar, analiticamente, os problemas identificados através de suas

características. Espera-se com isso possibilitar a descrição da situação real dos sistemas em análise para efeitos de comparação com as necessidades, desejos ou expectativas da comunidade. A caracterização da mobilidade, assim como outras análises, é altamente dependente de uma base de dado rica quantitativa e qualitativamente, diante disso a fonte de dados mais utilizada para obter-se informações de mobilidade de uma cidade são as pesquisas domiciliares de origem destino.

Destacam-se dois estudos que produziram esforços para caracterizar a mobilidade de Fortaleza, ambos utilizaram informações da pesquisa domiciliar de origem e destino da Região Metropolitana de Fortaleza (RMF) de 1996, que foram atualizados para 1999 durante o estudo de integração do Metrô de Fortaleza (METROFOR, 1999), são eles: Henrique (2004) que propôs e aplicou uma metodologia de análise espacial para caracterizar a mobilidade e a acessibilidade dos usuários do transporte público por ônibus e Sousa, Mesquita e Loureiro (2019) que caracterizou a evolução do padrão de mobilidade de Fortaleza entre os anos de 2000 e 2015.

Henrique (2004) propõe e aplica um método de caracterização da mobilidade e acessibilidade dos usuários do transporte público de Fortaleza, seu objetivo principal é identificar o padrão de distribuição dessas variáveis, assim como de outras variáveis socioeconômicas e de uso do solo, candidatas a explicarem fenômeno da autocorrelação espacial detectada na quantidade de deslocamentos realizados. Esse objetivo permitiu compreender as variações espaciais da mobilidade e identificar problemas de desigualdade espacial na mobilidade em Fortaleza.

Uma das etapas do método proposto por Henrique (2004) é exatamente a caracterização da mobilidade, nessa etapa é avaliado o comportamento espacial de duas variáveis: quantidade de viagens produzidas no transporte coletivo e quantidade de domicílios com renda até 5 salários-mínimos. As duas variáveis são escolhidas com propósito de identificar regiões de comportamento espacial semelhantes, distintos, bem como, avaliar o seu nível de relação.

Também se propõe a criação do seguinte indicador de taxa de mobilidade: viagens/domicílio, nesse indicador, o numerador representa as viagens produzidas em um dia típico pelos usuários do transporte coletivo, na hora de pico da manhã (6:30-7:30), enquanto o denominador expressa quais os domicílios cuja renda corresponde à classe social dos usuários cativos do sistema.

Os resultados de Henrique (2004) mostraram que, em Fortaleza, grande parte dos usuários cativos do sistema de transporte público pertencem às classes baixa e média e a maior parte dessa população reside em zonas periféricas. Henrique (2004) aponta também que a acessibilidade das zonas tem um impacto no nível de mobilidade desses usuários e que usuários em zonas mais periféricas possuem menos níveis de mobilidade, esses resultados demonstraram a existência de desigualdade espacial na mobilidade destes usuários.

Já a pesquisa de Sousa, Mesquita e Loureiro (2019), para caracterizar a evolução do padrão de mobilidade de Fortaleza, utiliza um software de modelagem integrada chamado *Tranus*. O *Tranus* é um modelo econométrico (IACONO *et al.*, 2008) que simula decisões locais e de viagens, sendo capaz de simular as intrarrelações dos subsistemas de atividades, uso do solo e transportes, considerando equilíbrio em cada um deles; modela também as inter-relações entre uso do solo e transportes, assim como o impacto do subsistema de atividades sobre os outros dois subsistemas urbanos (LOPES *et al.*, 2018).

Sousa, Mesquita e Loureiro (2019) também utilizaram informações da pesquisa domiciliar de origem e destino da Região Metropolitana de Fortaleza (RMF) de 1996 atualizados para o ano 1999, para calibrar os modelos de transporte do *Tranus* no ano base 2000.

Dessa pesquisa foram coletadas medidas de desempenho agregadas do sistema de transporte: divisão modal por tipo de viagem, percentual de deslocamentos por tipo e tempo médio de viagem. Para o ano de 2015, foram utilizados os parâmetros da base de 2000 atualizados e foram consideradas novas medidas de desempenho que foram obtidas de pesquisas de linha de contorno e de linha de travessia utilizadas durante o Projeto Fortaleza 2040 (PMF, 2015), também foram utilizados os dados de validação do transporte coletivo (ETUFOR, 2015) e a posse veicular por domicílios e tempos médios de deslocamentos obtidos pela Pesquisa Nacional de Amostra por Domicílios (PNAD) de 2015 (IBGE, 2015).

Sousa, Mesquita e Loureiro (2019) utilizaram os seguintes indicadores para avaliar a evolução entre os anos: Deslocamento na hora pico por modo; Distância; Custo (\$); Tempo de Viagem (Min); Tempo de espera (min). E, especialmente, foi avaliada a distribuição da origem e do destino das viagens nas zonas de fortalezas para os anos analisados.

A partir dos resultados das análises da evolução de cada um dos indicadores, Sousa, Mesquita e Loureiro (2019) avaliam que as condições da mobilidade em Fortaleza vêm agravando ao longo do tempo e que os indicadores de distância da viagem e tempo de espera apresentaram

um crescimento elevado, mostrando que a acessibilidade em média está pior. Também é percebido uma queda nos tempos de viagens, apontando para o investimento na rede de transporte público como principal causa dessa melhora. Esses resultados indicam um desequilíbrio entre os níveis de mobilidade e acessibilidade em Fortaleza.

Apesar desses resultados obtidos por Sousa, Mesquita e Loureiro (2019), percebe-se que existe um grande esforço para tornar os dados disponíveis viáveis de serem utilizados, além da necessidade de consideração de várias premissas no processo de modelagem da demanda. As etapas de calibração de parâmetros e expansão de amostras são etapas delicadas em que quaisquer alterações realizadas podem mudar os resultados obtidos o que torna os resultados bastante dependentes das premissas utilizadas.

Um dos aspectos que esses estudos têm em comum é a não consideração da heterogeneidade para a caracterização dessa mobilidade, ou seja, ambos os trabalhos consideram que a mobilidade dos usuários pode ser representada por indicadores únicos para cada zona. Com base no que foi discutido na seção 2.6, existem diferentes aspectos que impactam na mobilidade de cada usuário, fazendo com que essa mobilidade seja única, e, conseqüentemente, heterogênea dentro de um mesmo grupo, o que torna necessário, em caracterizações mais detalhadas, a consideração dessa heterogeneidade.

2.7 Padrão de mobilidade

Liang Liu *et al.* (2009) apontam que a compreensão da dinâmica dos habitantes e os seus padrões de mobilidade diários são essenciais para o planejamento e a gestão de equipamentos e serviços urbanos. O significado do termo padrão de mobilidade é pouco discutido na literatura e, portanto, há uma dificuldade de encontrar a definição do que é um padrão de mobilidade. Pode-se dizer que as análises de padrões de mobilidade têm propósitos parecidos com as análises de padrões de viagens. O termo padrões de viagens é comumente utilizado quando uma análise sobre um conjunto de viagens tenta revelar a sua regularidade e variação entre os indivíduos ao longo do tempo (KITAMURA, 2001).

Considerado os paradigmas do planejamento de transportes descrito no início deste capítulo, é possível dizer que o termo padrão de mobilidade é a evolução do termo padrão de

viagem, uma vez que o foco do planejamento era inicialmente garantir o tráfego a partir de infraestrutura, em seguida o foco se tornou as viagens e por último, o indivíduo.

Como não existe uma definição exata e para que não ocorra nenhuma confusão, neste estudo, um padrão de mobilidade é formado a partir da existência de uma ou mais características de mobilidade ou que possa afetar a mobilidade, de modo a representar uma população de interesse. Essas características serão responsáveis pela classificação e pelo agrupamento dos indivíduos, permitindo, assim, que esses indivíduos sejam caracterizados a partir do seu padrão de mobilidade.

A partir das seções anteriores, entende-se que cada indivíduo possui suas próprias características e que, portanto, é improvável que exista dois usuários com características completamente iguais, porém é comum que os usuários possuam uma ou mais características em comum. Assim, diante desta definição de padrão de mobilidade proposta, o agrupamento de indivíduos a partir de suas características em comum é o que permite a identificação dos padrões de mobilidade.

2.8 Conclusão e lacunas

Apresentado os diferentes entendimentos dos conceitos e as diferentes práticas na mobilidade e na acessibilidade é importante dizer que essas matérias se completam, durante o processo de planejamento urbano e que é necessário investigar melhor essa relação entre acessibilidade e mobilidade.

Nesse sentido, indagamo-nos se um usuário com maior acessibilidade, obrigatoriamente, terá maior mobilidade? Uma maior mobilidade implica diretamente uma maior acessibilidade? Melhor mobilidade e maior mobilidade significam as mesmas coisas? Quais os indicadores de mobilidade e acessibilidade podem ajudar a verificar essas hipóteses? Quais características de cada subsistemas que impactam nessa relação? São perguntas que ainda devem respondidas. Apesar de a relação entre mobilidade e acessibilidade ser uma lacuna importante, que o conhecimento científico deve preencher, não é objetivo deste trabalho investigar com profundidade esse assunto, mas levantar hipóteses e questionamentos para estudos futuros.

Em relação a caracterização da mobilidade, conforme já apresentado, em Fortaleza são encontrados dois métodos bem diferentes. O primeiro método indicado por Sousa, Mesquita e

Loureiro (2019) que só é possível com a utilização do software Tranus e do processo de calibração descrito no estudo. E o estudo de Henrique (2004), que propõe uma metodologia mais utilizada na literatura, possuindo uma melhor replicabilidade, além de ter o mesmo objeto de estudo desta pesquisa, que é o sistema de transporte público por ônibus em Fortaleza.

Henrique (2004) analisa o padrão de mobilidade de cada zona de Fortaleza a partir do indicador taxa de mobilidade (viagens produzidas por domicílios). Este tipo de análise considera a premissa de que os usuários residentes em uma mesma zona possuem o mesmo padrão de mobilidade. Porém, conforme apresentado neste capítulo, cada usuário é impactado pelas restrições de acessibilidade de forma singular, ou seja, a singularidade da relação de cada decisão de viagem e cada indivíduo a partir das restrições de acessibilidade vai contra a premissa utilizada por Henrique (2004).

Portanto, esta pesquisa se propõe a investigar a existência de uma heterogeneidade na mobilidade dos indivíduos em uma mesma zona de tráfego. O indicador de taxa de mobilidade proposto por Henrique (2004), parece adequado para analisar a mobilidade com uma visão mais equânime e sustentável do fenômeno. Por considerar, em seu cálculo, a quantidade de domicílios que possuam renda de até 5 salários-mínimos, o indicador é capaz de comparar a mobilidade considerando a situação social daquela zona.

Por exemplo, a afirmação de Henrique (2004) de que usuários residentes em zonas mais periféricas, por possuírem uma baixa acessibilidade, também possuem um baixo nível de mobilidade, só é possível pela utilização desse indicador, uma vez que a quantidade de viagens originadas nas zonas periféricas é relativamente maior as viagens geradas nas zonas mais centrais.

Os dados e o método utilizados por Henrique (2004) permitem que um retrato da situação da mobilidade de Fortaleza seja caracterizado, porém com a limitação da agregação por zona. Os resultados desta análise, devido à limitação da agregação por zona, podem ser insuficientes para o processo de planejamento, que tem como objetivo a mitigação da queda da demanda. A queda da demanda do sistema de transporte público por ônibus em Fortaleza evidenciada por Oliveira (2019) já era um problema levantado por Henrique (2004), porém a falta de dados longitudinais e desagregados não permitiam análises detalhadas sobre o problema específico.

Os indicadores utilizados por Sousa, Mesquita e Loureiro (2019) também parecem adequados, principalmente por representarem outros aspectos da mobilidade que não são

representados por Henrique (2004) como custo, tempo de viagem, tempo de espera, distância da viagem. Todavia diferente de Henrique (2004), o estudo de Sousa, Mesquita e Loureiro (2019) não se propõe avaliar esses indicadores por zona, portanto esses indicadores são apresentados como um valor médio para a cidade, conseqüentemente também indo contra a existência de uma heterogeneidade na mobilidade dos indivíduos.

A análise individual da demanda em um sistema de transporte pode ser um trabalho bastante complexo, por isso agregar indivíduos a partir de características em comum, como feito por Henrique (2004), Garcia (2016) e Sousa, Mesquita e Loureiro (2019), é frequentemente utilizado no planejamento de transportes. Essa dificuldade de analisar a demanda de cada indivíduo em um sistema de transporte, em que milhares de viagens ocorrem diariamente, é apontado, também, por Meyer e Miller (2001). Os autores sugerem, a fim de viabilizar a análise da demanda, a utilização de agrupamentos de indivíduos que possam ser agregados a partir de uma característica em comum.

Ortuzar (2011) também cita essa importância – para o planejamento adequado da demanda – da identificação e agregação de indivíduos em grupos que representem uma população de interesse, ou seja, é importante que no processo de compreensão dessa demanda seja identificado perfis dos usuários que utilizam o sistema através da utilização de características socioeconômicas e/ou padrões de deslocamento (diário, semanal e mensal), tornando esses grupos, assim, em facilitadores no processo de compreensão da demanda.

Portanto, entende-se que o processo de agregar usuários em grupos é benéfico e necessário para o planejamento da demanda. Porém, é importante definir um nível de agregação que permita alcançar os objetivos da análise e que, ao mesmo tempo, mostre sensibilidade para representar o fenômeno analisado.

Para superar essa limitação da dificuldade de captar a heterogeneidade dos padrões de mobilidade dentro de uma mesma zona, esta pesquisa irá separar os usuários do sistema de transporte público em grupos, que serão moldados a partir das características de mobilidade de cada indivíduo. Para isso, é premissa desta pesquisa, que os fatores a influenciar a mobilidade dos indivíduos influenciam de forma diferente cada um dos grupos.

Para que isso se torne possível, é necessária a utilização de dados com informações longitudinais das viagens desagregadas por indivíduo de uma grande amostra da população de usuários. Apesar da grande quantidade de dados analisados por Henrique (2004) e outros estudos,

a falta de dados longitudinais e desagregados é uma das principais barreiras para as análises mais detalhadas da demanda.

Para superar essa dificuldade, esta pesquisa propõe a utilização de dados de *Smart Card* do transporte público por ônibus de Fortaleza, base que possui informações das viagens realizadas pelos usuários do sistema. Esse tipo de dado, por possuir uma amostra quase populacional e com dados longitudinais, assegura uma riqueza de informações que não são encontradas nas bases de dados que normalmente são utilizadas no planejamento de transportes. Essas informações permitem uma melhor percepção da variabilidade espacial e temporal da mobilidade dos indivíduos, conseqüentemente, ajudam na investigação da hipótese da heterogeneidade dos padrões de mobilidade.

3 CARACTERIZAÇÃO E IDENTIFICAÇÃO DE PADRÕES DE MOBILIDADE A PARTIR DO *BIG DATA* DE TRANSPORTE PÚBLICO

3.1 *Big Data* de transporte público

O termo *Big Data* é definido pela National Science Foundation como bases de dados grandes, diversas, complexas e longitudinais que são geradas por instrumentos, sensores e transações na internet (NATIONAL SCIENCE FOUNDATION, 2011). Ainda não existe exatamente um consenso sobre o que é *Big Data*.

Kitchin (2013) cita características de *Big Data* que foram encontradas na literatura, como: Grande em volume; alto em velocidade: dados criados em tempo real; diverso em variedade; exaustivo no seu alcance: dados quase populacionais; alta resolução: o mais detalhista possível; relacionável por natureza: contém campos em comum que permitem a junção de várias bases; flexível: pode se estender tanto em atributos como em tamanho.

No planejamento de transportes, uma das mais conhecidas e utilizadas tecnologias são os sistemas de bilhetagem eletrônica (*Smart Card*). Esses sistemas permitem o usuário efetivar o embarque e pagar a tarifa da viagem através de um cartão, e são utilizados em muitos sistemas de transporte público urbano em todo o mundo (NUNES *et al.*, 2015).

Através dessa efetivação de embarque muitas informações daquela viagem podem ser obtidas a partir desse registro. Outro sistema utilizado é o de GPS dos ônibus, que permite a obtenção de dados complementares como a geolocalização das atividades registradas pelos sistemas de bilhetagem eletrônica. A geolocalização dos ônibus e, conseqüentemente, dos registros na base de *Smart Card* são importantes para alocação desses eventos no espaço, essas informações serão essenciais no processo de caracterização desses dados no espaço.

Além da bilhetagem eletrônica e do GPS, há uma outra fonte de dados chamada General Transit Feed Specification (GTFS). O GTFS é uma especificação de dados que tem como objetivo definir um formato comum para horários de transporte público e informações geográficas associadas. Os “feeds” do GTFS permitem que as agências de transporte público publiquem seus dados de trânsito e os desenvolvedores criem aplicativos que consumam esses dados.

Esse tipo de dado acaba sendo uma fonte rica de dados da oferta de diferentes sistemas de transportes, seja para analistas, planejadores, pesquisadores ou até usuários de sistemas de

transportes. Assim, podemos considerar que dados descritos podem ser considerados como *Big Data*. Essa definição permite entender que, por possuir uma maior quantidade de observações, o processamento e análise desses dados requer mais tempo e diferentes técnicas de análise, sendo necessária uma revisão sobre as melhores práticas da literatura.

A utilização dos dados oriundos do sistema de bilhetagem em conjunto com o sistema de GPS e o GTFS é cada vez mais reconhecida, seja pelo mundo acadêmico seja pelos órgãos e empresas responsáveis pela fiscalização e operação desses sistemas de transporte (PELLETIER *et al.*, 2011).

Os dados oriundos desses sistemas vêm sendo utilizados em diversos campos do planejamento de transportes, como análise dos padrões de viagens dos usuários de transportes público (MA, 2013; KIEU; BHASKAR; CHUNG, 2015; ZHAO; 2017), análise do comportamento (BLYTHE, 2017; ALI; KIM; LEE, 2016; BRIAND *et al.*; 2017), análise de intervenções e performance de transporte de ônibus (KIM; CHEON; LIM, 2011; PAU, 2014; JIN; JI; MOON, 2015; SMART; MILLER; TAYLOR, 2009), planejamento do sistema de transporte público (AUDOUIN; RAZAGHI; FINGER, 2015; GSCHWENDER; MUNIZAGA; SIMONETTI, 2016; UTSUNOMIYA; ATTANUCCI; WILSON, 2006).

Pelletier *et al.* (2011), em uma revisão de literatura sobre o uso de dados de *Smart Card* no transporte público, mostram que esses dados têm aplicações em qualquer nível de planejamento, seja em nível estratégico, tático seja operacional. No nível estratégico, os estudos buscam compreender o comportamento dos usuários que possam ajudar no planejamento a longo prazo. A grande quantidade de dados coletados fornece mais observações no espaço e tempo do que com qualquer outro meio de coleta de dados (PELLETIER *et al.*, 2011; AGARD *et al.*, 2006; BAGCHI; WHITE, 2005).

Como exemplo de análises em nível estratégico, podemos citar Agard *et al.* (2006), que tenta compreender o comportamento desses usuários, procurando hábitos de viagem semelhantes durante os dias da semana e durante as semanas. Trépanier e Morency (2010) calculam a vida útil do cartão e a sua relação com as características de uso dos usuários de bilhetagem eletrônica, as análises dos autores podem ajudar na modelagem da fidelidade dos usuários ao sistema e ajudar no planejamento de como o sistema pode ser mais atraente para reter mais usuários através da identificação de quais características são responsáveis por uma maior vida útil de cada cartão.

Zhou, Murphy e Long (2014) utilizam os dados de *Smart Card* em conjunto com uma pesquisa domiciliar para, respectivamente, comparar a eficiência na realização desses deslocamentos diários dos modos carro e transporte público e para realizar comparações entre os padrões de deslocamento diários de Beijing com cidades ocidentais. A combinação entre os dados teve objetivo de obter informações mais ricas sobre várias variáveis socioeconômicas, facilitando a compreensão sobre o comportamento de deslocamento de maneira mais concreta. O objetivo dos autores é que essas análises subsidiem o planejamento da cidade de Beijing a longo prazo.

A maior parte dos estudos que utilizam dados de *Smart Card* tem o objetivo de reconstruir viagens realizadas e identificar padrões temporais e espaciais para essas viagens. Na divisão proposta por Pelletier *et al.* (2011), esses estudos pertencem ao nível tático. Como exemplo podemos citar Bagchi e White (2005) que reconstroem as viagens e analisam os padrões de viagem a fim de realizar ajustes na oferta do serviço. Trépanier *et al.* (2007) e Munizaga *et al.* (2010) propõem análises que ajudam a descobrir o ponto de desembarque e origem, a fim de construir uma matriz origem e destino.

No nível operacional é possível encontrar estudos que utilizam dados de *Smart Card* para calcular indicadores precisos de desempenho em uma rede de transporte público, como aderência a horários, quilômetros de veículo e quilômetros de pessoas para cada corrida, rota ou dia (TRÉPANIÉ, 2009).

Existem estudos que também analisam características de pagamento e tarifas, para alcançar objetivos diferentes: identificar erros e inconsistências nas transações (CHAPLEAU; CHU, 2007); identificar oportunidades de melhorias nas políticas tarifárias (DEAKIN; KIM, 2001). Utsunomiya *et al.* (2006) e Seaborn *et al.* (2009) também buscam analisar mudanças nos horários programados com base nas quantidades de viagens diárias.

Halvorsen *et al.* (2019) apontam que a disponibilidade de dados de *Smart Card* facilita a avaliação de políticas de transportes em três níveis de agregação: Sistema (todos os passageiros), grupos de usuários e de forma individual a cada passageiro. O primeiro nível permite analisar os dados de passageiros agregados totalmente, ou seja, fluxos de linhas ou estações que são úteis do ponto de vista operacional.

Nesse caso, podem fornecer uma compreensão de como a demanda agregada mudou ao longo de determinado período. Para o segundo nível de agregação (grupos de usuários), os

autores defendem que as análises detalhadas dos comportamentos do usuário podem fornecer informações de como os impactos específicos de uma política atua em cada grupo.

Além disso, o monitoramento do comportamento de grupos específicos de usuários pode ajudar a entender quais características tornam as pessoas mais propensas a responder à intervenção e como as estratégias podem ser adaptadas para alcançar os diferentes comportamentos. Por fim, a agregação a nível individual pode fornecer uma ferramenta exclusiva para monitorar o comportamento do passageiro antes e depois da implementação de alguma intervenção. Com base na inferência de usuários que mudaram seu comportamento, modelos de escolha discreta são usados para identificar os principais fatores que contribuem para a decisão dos passageiros de mudar o comportamento e, portanto, informar projetos futuros mais eficientes.

Em Fortaleza, destacam-se três estudos importantes que já utilizaram esses dados: Freitas (2014) que buscou caracterizar essa demanda através da análise do retrato de um ano específico utilizando os dados de *Smart Card*, GPS e cadastro do bilhete único; Braga (2019) que utilizou os dados de *Smart Card*, GPS e GTFS para analisar a variabilidade de indicadores de acessibilidade por transporte público às oportunidades de trabalho e educação; Oliveira (2019) que analisou a evolução da demanda de transporte coletivo por ônibus utilizando dados de todo o *Big Data* do transporte público de Fortaleza.

Freitas (2014) utiliza os dados oriundos do sistema de bilhetagem e dos dados de GPS instalado na frota de ônibus, para caracterizar a problemática do sistema de transporte público de Fortaleza olhando para um recorte no tempo desses dados. O autor avançou na utilização desses dados para descrever a demanda do sistema e demonstrou a importância e o potencial da utilização desses dados.

Para identificar a origem das viagens analisadas, Freitas (2014) propôs a utilização do local da primeira validação do dia como local de origem. Entende-se como validação o registro da presença do usuário no sistema por meio do cartão eletrônico ou o momento em que o usuário utiliza o seu cartão para pagar a tarifa de utilização do sistema. Como método de validação da premissa de que o local da primeira validação é próximo a residência do usuário, realiza uma análise de comparação dessas localizações. Utilizando zoneamento por bairros de Fortaleza como nível de agregação, ele encontrou que 66% dos usuários realizam a validação em zona diferente da zona de residência. Essa evidência de que as validações são realizadas longe das residências

não são reconhecidas por Freitas (2014) como uma limitação, o que traz impactos negativos para as suas análises.

Braga (2019) também utiliza dados de *Smart Card* para analisar a variabilidade de indicadores de acessibilidade por transporte público para oportunidades de trabalho e para educação. Braga (2019) propõe também uma metodologia de comparação para o cálculo dos indicadores de acessibilidade.

Para realizar essa comparação ele propõe utilizar os dados planejados da oferta do sistema, que estão no formato GTFS (Especificação Geral de Feed de Trânsito) e uma correção desses dados a partir da utilização das informações de *Smart Card*. Essa metodologia permitiu constatar o efeito da variabilidade dos tempos de viagem na rede para a acessibilidade dos usuários do sistema. Assim, Braga (2019) produziu um avanço no tratamento desses dados e permitiu a replicabilidade do método de georreferenciar os registros das validações da base de dados de *Smart Card*.

Por fim, Oliveira (2019) propõe analisar a evolução da demanda do transporte coletivo por ônibus em Fortaleza através do uso de *Big Data*. Em seu estudo, Oliveira (2019) faz uma revisão de literatura sobre o uso dos dados de *Smart Card* no planejamento de transportes. Além disso, a fim de agrupar usuários que possuam características e padrões de mobilidade em comum foi aplicado um método de clusterização bastante utilizado na literatura que apresentou resultados para indicar que o método aplicado pode ajudar no processo de agrupamento de usuário.

Esse método permitiu identificar grupos formados por usuários com características diferentes. Foram identificados e interpretados 10 diferentes grupos a partir da utilização de 7 atributos responsáveis por representar as características e os padrões de mobilidade dos usuários. Oliveira (2019) reconhece que a combinação de muitos grupos e a utilização de muitos atributos dificulta, por vezes, a interpretação dos agrupamentos obtidos e que a inserção de atributos como o período do dia de utilização do transporte coletivo aumenta significativamente a quantidade de grupos estabelecidos. Também é reconhecido por Oliveira (2019) que a associação de características socioeconômicas pode facilitar no processo de interpretação desses grupos. Apesar das dificuldades em interpretar os diferentes grupos encontrados, reconhece-se que os resultados obtidos podem ajudar a análise da demanda e da mobilidade e, portanto, esses resultados são motivadores e serão base para esta pesquisa.

Pelletier *et al.* (2011) apontam que estudos comportamentais, no âmbito de usuários, são difíceis por causa do anonimato do cartão e pela ausência de dados socioeconômicos do usuário. E, como apontado por Freitas (2014) e Oliveira (2019), os dados de Fortaleza possuem uma realidade diferente de alguns casos visto na literatura. Para Fortaleza, existe disponível um conjunto de dados de cadastro para os usuários do sistema do Bilhete Único Fortaleza e esses dados trazem algumas características socioeconômicas dos usuários, permitindo o aprofundamento em análises de demanda.

3.2 Identificação de padrões de mobilidade a partir de dados de *Smart Card*

Nesta seção discute-se o propósito da identificação de padrões de mobilidade, estudos relevantes que identificam e analisam padrões de mobilidade, utilizando dados de *Smart Card*, bem como, quais os padrões encontrados e esperados por esses estudos. Serão apresentados também quais métodos e ferramentas vêm sendo mais utilizados para identificar padrões de mobilidade através de dados de *Smart Card*.

Conforme apresentado no capítulo anterior, nesta pesquisa, um padrão de mobilidade é formado a partir da existência de uma ou mais características de mobilidade ou que afete a mobilidade que possa representar uma população de interesse. Na literatura, a identificação de padrões de mobilidade a partir de dados de *Smart Card* é realizada, principalmente, a partir da criação de perfis para cada usuário.

Esse perfil é construído a partir das características dos usuários, que podem ser características socioeconômicas ou características das viagens. A partir da criação desses perfis, são aplicadas ferramentas de clusterização (*clustering*) que agregam os usuários com características em comum e conseqüentemente separam os que possuem características diferentes (AGARD *et al.*, 2006; MORENCY, 2006; KIM; KIM; SOHN, 2018; AGARD; TRÉPANIÉ, 2013; EL MAHRSI *et al.*, 2014; EL MAHRSI *et al.*, 2017; PIERONI, 2018; OLIVEIRA, 2019).

Em sua essência, a identificação de padrões busca separar os padrões que são regulares ou típicos de uso do sistema, ou padrões que sejam esporádicos ou atípicos, porém tudo depende das características de cada indivíduo que são escolhidas para construir os perfis, portanto, se é utilizado uma característica socioeconômica para construir os perfis dos usuários, estes serão classificados considerando essa característica. Essa classificação, seja de usuários seja do mercado,

é apontada por Faroqi *et al.* (2019) como um facilitador para o desenvolvimento de políticas específicas para cada grupo identificado. Halvorsen *et al.* (2019) aponta também que políticas públicas futuras podem ser melhoradas se a heterogeneidade do usuário for considerada através da identificação de grupos.

Agard *et al.* (2006), em um estudo que objetiva caracterizar o comportamento do usuário, a partir da identificação de grupos de utilizadores com hábitos de viagem semelhantes, a pesquisa propõe e aplica um método que combina duas ferramentas de clusterização. O método proposto utiliza todo o conjunto de dados e divide-o em grandes grupos homogêneos com base nos padrões de viagem observados diária e semanalmente, a fim de identificar agrupamentos naturais de usuários que possuam uma utilização do sistema semelhante durante os dias da semana e durante as semanas. Para identificar os agrupamentos são utilizadas 20 variáveis binárias para construir os perfis, onde cada variável representa um período de um dia analisado (4 períodos x 5 dias da semana), construindo assim um padrão temporal de utilização do sistema.

As características de data e hora da validação da viagem também são utilizadas por Morency *et al.* (2006) para construir os perfis temporais. O objetivo principal desse estudo é detectar padrões típicos e a regularidade de viagem dos perfis construídos. Para compreender melhor as variáveis investigadas os autores realizaram uma análise descritiva antes de aplicar a técnica de clusterização (*k-means*) de alguns indicadores como número de embarques classificados por tipo de cartão. De acordo com os autores, a análise descritiva foi essencial para dar início as etapas de clusterização, pois os resultados permitiram uma melhor definição dos parâmetros necessários requeridos pelas ferramentas de clusterização.

Ma *et al.* (2013) procuram identificar os padrões temporais e espaciais das validações de cada usuário e utilizam as seguintes variáveis para construir os perfis dos usuários: número de dias de uso, número de embarques em horas similares, número de rotas similares, número de utilização de paradas similares.

Nesse caso, a técnica de clusterização utilizada foi uma adaptação do *k-means* chamada *k-means++*. Algumas restrições para a identificação dos padrões foram utilizadas, um usuário só foi considerado, caso ele possuísse pelo menos três registros, em dias distintos, que tivessem sido realizadas dentro de um mesmo período durante um dia típico em uma distância variando em até 200m. O objetivo dessas restrições foi remover os usuários que utilizavam o sistema de forma esporádica. Além disso, uma outra ferramenta de clusterização chamada Density-based spatial

clustering of applications with noise (DBSCAN) foi utilizado. O DBSCAN é uma ferramenta de clusterização normalmente utilizada para encontrar padrões espaciais. O autor utiliza com o objetivo de identificar de dias atípicos de uso a partir da localização dos embarques e, assim, desconsiderá-los das análises.

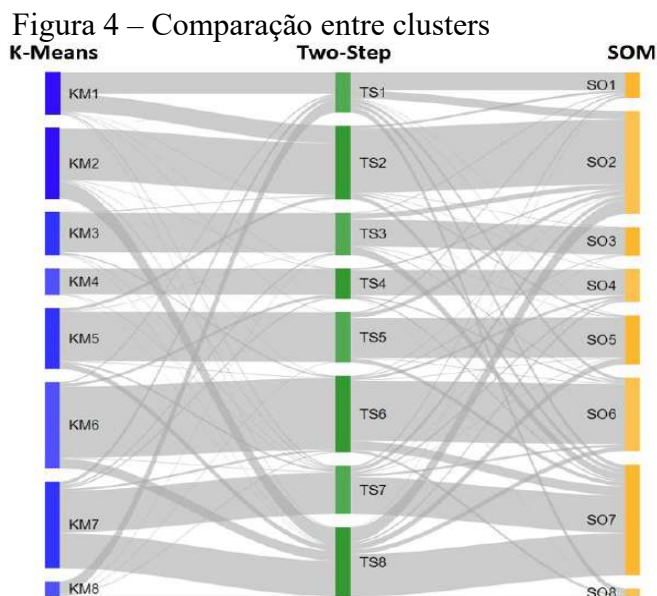
El Mahrsi *et al.* (2014, 2017) utilizam dados de *Smart Card* do sistema de transporte público da cidade de Rennes na França para identificar esses grupos através de características das viagens dos usuários. El Mahrsi *et al.* (2014, 2017) constrói perfis temporais de passageiros com base em informações de embarques realizados durante um período de 23 dias e, então, a partir da aplicação de uma ferramenta de clusterização 16 clusters são identificados. Esses cluster então são utilizados para caracterizar e avaliar como os hábitos de viagem (hora e dia da semana de utilização) dos passageiros se relacionam com as características socioeconômicas (faixa etária e ocupação). Para isso, foi utilizado uma base de dados com características socioeconômicas que permitiram estimar as características socioeconômicas dos passageiros a partir da origem do seu embarque.

Pieroni (2018) utilizou os dados de *Smart Card* e uma base de dados socioeconômicos para investigar os padrões de viagens espaço-temporais de vários dias de usuários de transporte público que residem em assentamentos precários na cidade de São Paulo. Foram calculados indicadores para representar as características individuais e de viagem de cada usuário: hora da primeira validação, frequência semanal, distância espacial entre validações, distância temporal entre validações, desvio padrão da primeira validação, desvio padrão da distância temporal, desvio padrão da distância espacial das validações, frequência, renda média e características de uso do solo da região de embarque. Essas características avançam em relação a outros estudos, pois a combinação dessas características permitiu uma criação dos perfis de usuário que considera aspectos importantes para investigação de padrões de mobilidade.

A partir de uma revisão de literatura, Pieroni (2018) optou por realizar a identificação dos padrões a partir de três diferentes ferramentas de clusterização: *K-means*, SPSS TwoStep e SOM. O SPSS TwoStep é um método proposto pela IBM (2011) que combina dois processos de clusterização, o propósito desse método é identificar uma quantidade inicial de clusters a partir de um método não hierárquico, e então, aplicar um método de clusterização hierárquica, desse modo esse método funciona bem para grandes bases de dados.

Self-organizing map (SOM) é uma rede neural não supervisionada usada para visualizar conjuntos de dados de alta dimensão em representações dimensionais inferiores, denominada mapa (KOHONEN, 2001; LYNN, 2018). Alguns estudos usam o SOM para organizar dados demográficos e habitacionais coletados de censos e pesquisas para investigar padrões de mudança nos sistemas urbanos e regionais (HATZICHRISTOS, 2004; SKUPIN; HAGELMAN, 2005). Pieroni (2018) também utiliza a ferramenta DBSCAN para a identificação de clusters espaciais das validações, permitindo, assim, não só estimar pontos de embarque médio que possam representar a origem dos usuários como identificar validações atípicas de cada usuário.

Uma vez identificado os grupos, Pieroni (2018) utiliza um diagrama de Sankey (Figura 4) para avaliar e comparar a estabilidade destes pelas três ferramentas. O Diagrama de Sankey é uma representação visual de um fluxo, a sua utilização é facilitar o senso de interpretação desse tipo de informação, o fluxo envolve a transferência de algum objeto em análise de uma etapa para outra, essas transferências são representadas por linhas cujas espessuras são proporcionais ao volume do objeto que está sendo transferido. Pieroni (2018) conclui que os fluxos principais visualizados na Figura 4 apresentam maioridade dentre as variações.



Fonte: Pieroni (2019).

Existem poucos estudos que utilizam dados de *Smart Card* e que testam e comparam os resultados da aplicação de diferentes ferramentas de clusterização como faz Pieroni (2018).

Porém, o autor não utiliza os resultados do diagrama de Sankey para sugerir qual dentre as ferramentas é a melhor para analisar o fenômeno em questão, pelo contrário, ele utiliza os resultados das três ferramentas para fortalecer o embasamento de que os grupos identificados e os resultados são servem para o propósito de análise dados de *Smart Card*, uma vez que cada um dos métodos possui diferentes critérios e premissas de uso e que, mesmo assim, os resultados foram consistentes para os três, com clusters semelhantes formados.

Para indicar qual o melhor método de clusterização é utilizado o indicador S_Dbw, que é também utilizado e sugerido por Lie *et al.* (2010). O indicador S_Dbw leva em consideração a variância intra-cluster e a densidade inter-cluster. O método *k-means* apresentou o melhor resultado para o indicador S_Dbw para a base de dados utilizada por Pieroni (2018).

Diferentes padrões de mobilidade podem ser definidos conforme às características individuais e de viagem escolhidas para construir os perfis dos usuários. Como é apresentado, a literatura traz esses perfis de usuários sendo construídos a partir da combinação de alguns atributos. Normalmente, são escolhidos atributos que representam a utilização do sistema (BRIAND *et al.*, 2015; HUANG *et al.*, 2015; MA *et al.*, 2013) e/ou atributos que representam características do uso do solo e da atividade dos usuários (PIERONI, 2018). A literatura, então, demonstra uma maior utilização de variáveis que representam as características de viagens dos usuários do que de variáveis representativas das características individuais, isso se dá principalmente pela limitação dos tipos de dados de *Smart Card*, os quais são dados de viagens realizadas e que, normalmente não possuem muitas informações socioeconômicas dos usuários.

3.3 Ferramentas de clusterização (*Clustering*)

Conforme visto na seção anterior, a utilização do *Big Data* de transporte público como fonte de informação para o planejamento de transporte é, frequentemente, encontrada na literatura. El Mahrsi *et al.* (2017) apontam que esses dados possuem observações mais completas, mais precisas e que eles podem ser rastreáveis a um nível individual e que essas características trazem alguns desafios para as análises a serem realizadas. Em seguida, o autor sugere a utilização de técnicas de clusterização como ferramentas essenciais no processo de extração de mais conhecimento (comportamento de grupos, padrões frequentes etc.) dessas bases de dados.

Técnicas de clusterização (*clustering*) são, muitas vezes, utilizadas para a identificação de padrões de mobilidade a partir do agrupamento de usuários que possuem características semelhantes.

A Clusterização é uma ferramenta de data mining (Mineração de dados). A mineração de dados é descrita por Witten e Frank (2005) como uma forma de extrair informações implícitas, anteriormente, desconhecidas e, potencialmente, úteis dos dados. Também pode ser considerado um termo mais amplo de um campo interdisciplinar composto por análises estatísticas, sistemas de banco de dados, aprendizado de máquina, reconhecimento de padrões, redes neurais, sistemas fuzzy e outras técnicas de computação (VELICKOV; SOLOMATINE, 2000).

Essa ferramenta é, normalmente, utilizada para encontrar grupos de objetos que são semelhantes entre si e diferentes dos objetos de outro grupo, é utilizada, muitas vezes, com o propósito de realizar análises exploratórias de padrões em diversas áreas do conhecimento (JAIN *et al.*, 1999).

De acordo com Pinto-Correia *et al.* (2006), as análises de clusters (*clustering*), com base em combinações de indicadores são ferramentas, constantemente, utilizadas em análises para identificar grupos que possuem indivíduos com comportamento semelhante. Morency *et al.* (2006) também propõem e utilizam ferramentas de mineração de dados para a identificação de padrões e para extrair conhecimento de grandes bancos de dados.

A partir da revisão de literatura descrita na seção anterior, percebe-se que o *k-means* e o DBSCAN são as ferramentas de clusterização mais comuns em análises de dados de *Smart Card* com objetivo de identificar padrões de mobilidade e grupos de usuários, porém outras ferramentas como Hierarchical Ascending Clustering (HAC), SPSS TwoStep e SOM também são utilizadas, iremos detalhar um pouco de algumas delas a seguir.

3.3.1 Density-based spatial clustering of applications with noise (DBSCAN)

O DBSCAN, inicialmente, proposto por Ester *et al.* (1996), fornece uma solução flexível para a análise de padrões de viagens espaciais (KIEU; BHASKAR; CHUNG, 2015; MA *et al.*, 2013). A principal desvantagem desse método é que ele é muito sensível aos parâmetros de entrada. Algumas vantagens dessa ferramenta são listadas por Kieu, Bhaskar e Chung (2015):

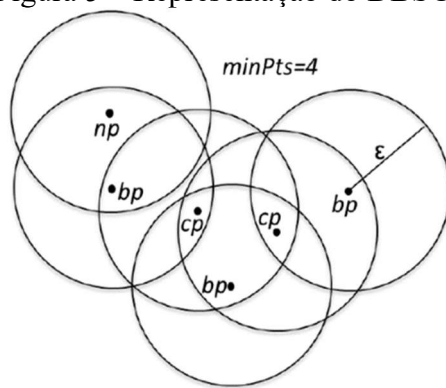
- i) o DBSCAN identifica grupos de alta densidade e ruídos que não pertencem a nenhum grupo. Na análise do padrão de viagem, o ruído é um padrão de anomalia

- e os *clusters* são o padrão de viagem espacial regular;
- ii) o *DBSCAN* identifica um *cluster* de qualquer formato e tamanho. Normalmente, em análises espaciais de padrão de deslocamento, os *clusters* identificados podem possuir qualquer forma e tamanho;
 - iii) o *DBSCAN* não requer predeterminação de núcleos iniciais ou número de *clusters*. Esse recurso também é muito importante para análise de padrões de viagem porque o número de padrões de um passageiro individual é desconhecido.

No *DBSCAN*, os pontos são agrupados por regiões de alta densidade, enquanto os pontos localizados em regiões de baixa densidade são caracterizados como *outliers*. Para identificar regiões de alta e baixa densidade, o algoritmo *DBSCAN* requer dois parâmetros como dados de entrada para seu processamento: distância ϵ e o número mínimo de pontos para definir um *cluster*, **minPts**.

Com base em um conjunto de pontos, o *DBSCAN* agrupa pontos próximos uns dos outros com base em uma medição de distância. E funciona basicamente da seguinte maneira: Um ponto é classificado como **cp** (*central point*) se houver pelo menos uma quantidade de **minPts** pontos com uma distância inferior do que ϵ a ele. É classificado como um **bp** (*board point*) se estiver a menos de ϵ de distância de um **cp**, mas não possuir uma quantidade de **minPts** com uma distância inferior do que ϵ a ele. Um **np** (*noise point*) são todos os outros pontos que não se encaixam como um **cp** ou **bp** (Figura 5).

Figura 5 – Representação do DBSCAN



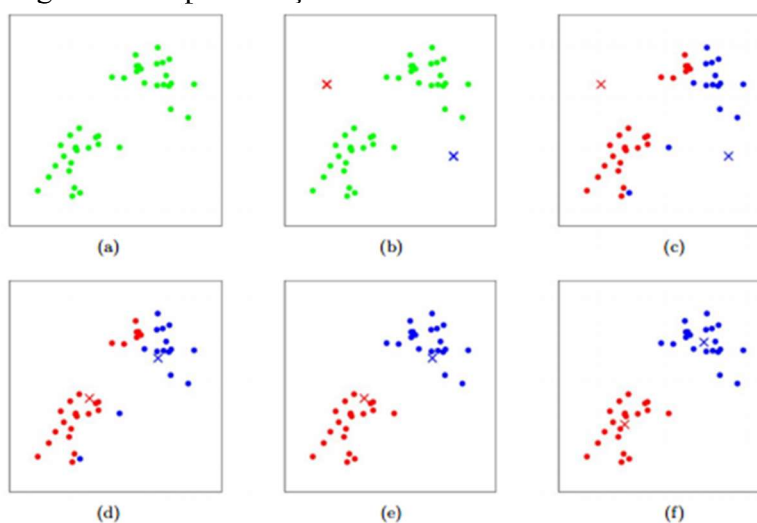
Fonte: Kieu, Bhaskar e Chung (2015).

3.3.2 *K-means*

O *k-means* é um dos algoritmos de clusterização mais simples e mais utilizados para realizar a identificação de clusters com base em atributos. É um método numérico, iterativo e não determinístico, ou seja, o algoritmo pode fornecer diferentes resultados se o mesmo for executado várias vezes com os mesmos dados, e tem como objetivo classificar dados através da identificação de grupos (NAMRATHA; PRAJWALA, 2012).

Na Figura 6 é possível compreender o passo a passo do funcionamento do algoritmo. Supõe-se um conjunto de dados distribuídos no espaço que pode ser multidimensional (a) e assumindo que existem 2 grupos, 2 centroides aleatórios são gerados (b) e então, calcula-se a distância de cada ponto até os dois centroides. O ponto será atribuído ao centroide mais próximo e, portanto, ao grupo mais próximo (c). A partir da primeira definição dos grupos, calcula-se novamente o centroide (agora de um grupo já definido) e o processo se repete iterativamente quanto se queira (d), (e) e (f).

O *k-means* usa uma função que busca minimizar a distância quadrática dos pontos aos centroides dos clusters (ZAKI; MEIRA, 2014). Os algoritmos de clusterização usam essa distância para separar os objetos em grupos diferentes, inclusive, a soma dessas distâncias dentro e fora dos clusters é, frequentemente, utilizada como indicador de qualidade para clusters. Para executar o *k-means* é necessário definir como parâmetro de entrada o número de agrupamentos esperado, o que pode ser uma desvantagem caso o pesquisador não saiba quantos grupos queira encontrar. Outra desvantagem é que o algoritmo começa através de uma definição aleatória de centroides pode levar a diferentes resultados para uma mesma base de dados e a uma minimização local dos erros quadráticos e não global.

Figura 6 – Representação do *k-means*

Fonte: Chris Piech (Stanford CS221).

Como solução, Morency (2007) e Pieroni (2018) sugerem a utilização de um algoritmo que permite diversas simulações a partir de diferentes pontos de partidas, ou seja, executar n vezes o *k-means* e identificar qual a melhor saída dentre os n resultados obtidos e definir como parâmetro um alto número de interação. Esse método possibilita o algoritmo identificar quais centroides geram clusters com a menor soma da distância dos quadrados entre as observações. Pieroni (2018), em seu método, executa o *k-means* 10 vezes e realiza 300 interações para cada execução.

3.3.2.1 Número de grupos (*k-means*)

Diferente dos outros algoritmos discutidos nessa revisão, o *k-means* necessita, como parâmetro de entrada, da quantidade de grupos (clusters) a serem identificados. A literatura traz três caminhos para definição de grupos, que podem ser realizadas por métodos estatísticos conforme feito por Zhao *et al.* (2017) e Pieroni (2018), através da escolha do analista a partir de sua experiência, conhecimento do fenômeno e objetivos da análise (MORENCY *et al.*, 2007), e por último, através de uma análise descritiva simples e espacial dos dados (AGARD; TRÉPANIÉ, 2013). Por exemplo, a análise descritiva para o problema de caracterização da mobilidade pode ajudar no processo de entendimento das características de heterogeneidade da demanda.

Os métodos estatísticos da literatura propõem o cálculo de indicadores para estimar o número ótimo de *clusters*, como coeficiente o *average silhouette*, que é uma medida de

separabilidade e compacidade dos grupos (ZHAO *et al.*, 2017) ou o método denominado *gap statistic*, proposto por (FRIEDMAN *et al.*, 2001), que realiza simulação de Monte Carlo e compara uma dada formação de grupos com outra formação totalmente aleatória. A ideia é comparar a dispersão dentro do *cluster* com sua expectativa sob uma distribuição de referência nula apropriada (TIBSHIRANI *et al.*, 2001).

Outro método que não é aplicado na literatura de *Smart Card*, mas é facilmente encontrado na literatura de clusterização em geral, chama-se *elbow method*. É um método visual, considerado um dos mais antigos para determinar o número de *clusters*, ideal para um conjunto de dados (KODINARIYA; MAKWANA, 2013). A ideia é que seja realizado a clusterização a partir do número de grupos (K) igual a 2 e vá aumentando a cada passo em 1. Em cada passo é identificado os *clusters* e suas variações. Espera-se que com algum valor para K, a variação dos *clusters* caia drasticamente e, depois disso, atinja uma estabilidade em que não apresente grande alteração para valores maiores de K. Assim, quando a estabilidade da variação for notada, o número de grupos para essa variação é o valor K que você deseja. A intuição é que ao aumentar o número de *clusters* irá, naturalmente, diminuir a variação, mas que, em algum ponto, a diminuição dessa variação é saturada, e o efeito visual do *elbow method* (Método do cotovelo) reflete isso.

3.3.3 Hierarchical Ascending Clustering (HAC)

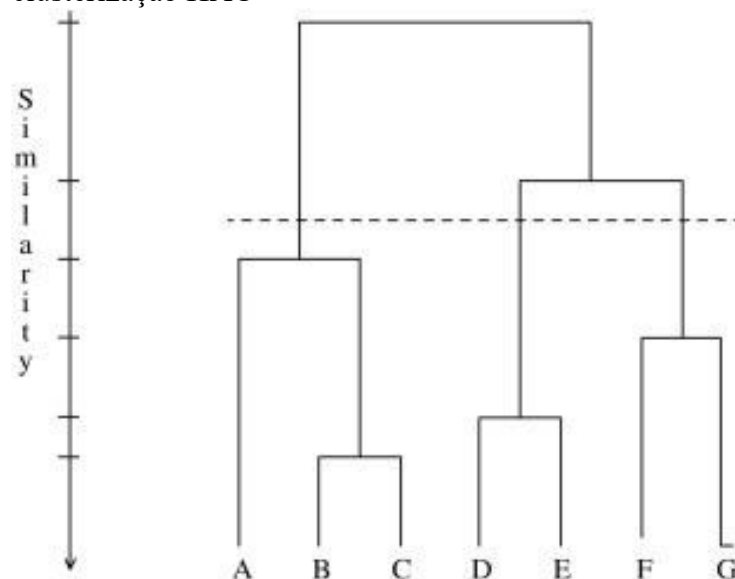
O *Hierarchical Ascending Clustering* (HAC) é uma ferramenta de clusterização que busca construir uma hierarquia entre os *clusters* identificados. As estratégias de agrupamento hierárquico, geralmente, podem ser do tipo aglomerativo ou do tipo divisivo (ROKACH *et al.*, 2005). No tipo aglomerativo, cada observação começa em seu próprio *cluster* e pares de *clusters* são mesclados à medida que se sobe na hierarquia, chamada de abordagem “de baixo para cima”. Já para o tipo divisivo, todas as observações começam em um *cluster* e as divisões são realizadas recursivamente à medida que se desce na hierarquia, abordagem chamada “de cima para baixo”.

Bhagat *et al.* (2016) listam como as principais vantagens do método de clusterização HAC a não necessidade de definição de número de *cluster* com antecedência, os resultados que apresentam toda uma hierarquia de *cluster*, que são apresentados a partir de um dendrograma, o que garante uma compreensível visualização dos resultados.

Além disso, o fato de existir uma hierarquia de clusters garante uma flexibilização do nível de granularidade dos clusters, que, basicamente, significa o tamanho de clusters. Como desvantagens do método, Bhagat *et al.* (2016) citam o fato de quando os clusters hierárquicos são mesclados ou divididos não podem mais ser desfeitos, além de necessitar como entrada uma matriz de distância de similaridade o que implica que este tipo de algoritmo não é indicado para grandes bases de dados.

A representação em dendrograma dos resultados de um algoritmo de agrupamento aglomerativo hierárquico pode ser visualizada na Figura 7. A linha tracejada serve como ponto de parada desse algoritmo específico e produz três *clusters*. Um *cluster* contém A, B e C; outro contém D e E e o *cluster* final contém F e G, apresentando um algoritmo com um critério de similaridade mais restrito resultaria em menos *clusters* no ponto de parada. Da mesma forma, um critério mais flexível resultaria em mais *clusters* (AKMAN *et al.*, 2019).

Figura 7 – Representação dos resultados de uma clusterização HAC



Fonte: Akman *et al.* (2019).

3.3.4 Distância de similaridade

Os algoritmos de clusterização expostos nas seções anteriores utilizam as distâncias de similaridade para agrupar observações de uma base de dados que são semelhantes nos mesmos

clusters, enquanto observações diferentes ou distantes são colocadas em clusters distintos (AGHABOZORGI *et al.*, 2018). A similaridade é uma quantidade que reflete a força do relacionamento entre dois itens de dados, ela representa quão semelhantes são as duas observações de dados a partir dos atributos escolhidos para representar essa observação.

Essa medida de similaridade é, normalmente, calculada a partir de funções de distância, como distância Euclidiana, distância de Manhattan, distância de Minkowski, similaridade por Cosseno etc. (IRANI; PISE; PHATAK, 2016). Essas distâncias são baseadas em medidas geométricas, porém, há também percursos baseados em estatísticas, que buscam medir distância entre as observações a partir de sua correlação, são exemplo delas o coeficiente de correlação de Pearson e o coeficiente de correlação de Spearman (IWASAKI; KUSNE; TAKEUCHI, 2017).

A forma de representação dessas distâncias é através de uma matriz simétrica ($n \times n$), em que n é a quantidade de observações de um conjunto de dados. Essa matriz é preenchida com as distâncias que são calculadas usando uma das medidas listadas anteriormente. Essa matriz se chama matriz de similaridade ou, se a distância for usada como medida de dessimilaridade, a matriz de distâncias (VANDEGINSTE *et al.*, 1998).

Na literatura que utiliza dados de *Smart Card*, a distância mais utilizada é a Euclidiana. Porém, são poucos os estudos que discutem qual medida de similaridade deve ser utilizada e quais os impactos da utilização de diferentes medidas nos resultados dos clusters. Agard e Trépanier (2013) avançam na discussão sobre as medidas de similaridade utilizadas e concluem que, para determinados casos, as medidas de similaridade de distâncias Euclidiana e Manhattan podem não ser uma boa opção para o método de análise proposto por eles.

Em sua metodologia, Agard e Trépanier (2013) precisam que usuários que validam em horários espaçados, mas diferentes, tenham níveis diferentes de similaridade. Por exemplo, a similaridade de um usuário que valida entre 07:00h e 08:00h e um usuário que valida entre 08:00h e 09:00h deve ser diferente da similaridade de um usuário que valida entre 07:00h e 08:00h e um usuário que valida entre 09:00h e 10:00h. No método proposto, as distâncias Euclidiana, Manhattan e Hamming não são capazes de diferenciar a similaridade desses usuários. Como forma de resolver esse problema, é sugerido pelos autores a transformação da distância euclidiana para coordenada polar e a incorporação de outras variáveis no modelo, como a frequência de uso, além de uma medida de distância temporal entre os usos. A coordenada polar considera que cada ponto no plano é determinado por uma distância e um ângulo em relação a um ponto fixo de referência.

Percebe-se que os tipos de atributos escolhidos influenciam, diretamente, em qual distância de similaridade deve ser utilizada. Por exemplo, as variações entre atributos categóricos (qualitativos) podem não ser percebidas por distâncias de similaridade que tem como base um cálculo geométrico. Já para atributos quantitativos, as variações entre eles são facilmente percebidas por distância de similaridade baseadas em medidas geométricas ou baseadas em estatísticas. Portanto, a medida de similaridade a ser escolhida depende dos atributos selecionados para construir os perfis de usuários e gerar os *clusters*.

3.3.5 Atributos (parâmetros de entrada)

Outro tópico importante sobre a clusterização em dados de *Smart Card* é sobre os atributos que servem como parâmetro de entrada para técnicas de clusterização. Jain *et al.* (1999) apontam que é necessário selecionar variáveis ou atributos que melhor representem o fenômeno a ser estudado, podendo ser esses atributos quantitativos ou qualitativos.

Nesse sentido, as distâncias de similaridade são calculadas a partir dos valores desses atributos, por tanto a escolha deles é essencial nesse processo. Existem três aspectos importantes sobre os atributos a serem discutidos: quais atributos, como comparar atributos que possuem escalas e variações diferentes e os pesos que cada atributo terá na formação dos clusters. Essas características serão discutidas posteriormente.

Conforme exposto nas seções anteriores, a identificação de padrões de mobilidade a partir de clusterização necessita dos perfis de cada usuário que são construídos através de atributos, que representam as características de viagens dos usuários e/ou aspectos socioeconômicos (Tabela 1).

Tabela 1 – Atributos encontrados na literatura de clusterização a partir de dados de *Smart Card*

Autor	Atributos	Dados	Distância de similaridade	Método
Agard <i>et al.</i> (2006)	4 períodos do dia x 5 dias da semana (20 variáveis binárias)	5 semanas	Não especificada	<i>HAC / k-means</i>
Morency (2006)	Hora do dia de utilização e dia da semana	277 dias	Não especificada	<i>k-means</i>
Agard e Trépanier (2013)	Frequência, hora do dia, variabilidade da hora de uso, distância temporal entre usos	1 ano	Distância euclidiana	<i>k-means</i>
Ma <i>et al.</i> (2013)	Número de dias de uso, número de embarques em horas similares, número de rotas similares, número de utilização de paradas similares	Semana típica	Distância Euclidiana	<i>k-means</i> ++
Kieu <i>et al.</i> (2015)	DBSCAN - Localização da validação k-means - Minuto de utilização	4 meses	Não especificada	<i>DBSCAN</i>
Pironi (2018)	DBSCAN -Localização da validação k-means -Hora da primeira validação, frequência semanal, distância especial e temporal entre validações, desvio padrão da primeira validação e da distância temporal e especial das validações, frequência, renda média, uso do solo	11 semanas	Distância euclidiana	<i>DBSCAN / k-means</i>

Fonte: Adaptado de Pironi (2018) e Oliveira (2019).

Em geral, as características de usuários como hora de utilização são amplamente utilizadas pelos autores, além da frequência de utilização e da localização das validações/embarques. O intervalo temporal entre usos também é usado em alguns estudos, podendo ser relacionado à duração das atividades, as características de viagens são obtidas normalmente através dos dados de *Smart Card* e as características socioeconômicas são obtidas através de bases de dados complementares (Pesquisa domiciliar e Bases de cadastros). Esses atributos são essenciais para moldar e interpretar os clusters encontrados em cada um dos estudos listados.

Uma vez que os atributos são escolhidos, é necessário que eles estejam em uma mesma escala, caso contrário as distâncias de similaridade serão baseadas em atributos com formatos, escalas e variações diferentes. Esse processo se chama padronização e trata-se de garantir que os dados sejam, internamente, consistentes, ou seja, fazer que esses dados tenham o mesmo conteúdo e formato. A padronização é essencial em algoritmos de agrupamento com conjunto de dados, contendo diversos tipos de variáveis, pois eles podem gerar um resultado inesperado, dependendo da distribuição das observações entre as variáveis (BARROSO; ARTES, 2003).

Berthold e Höppner (2016) em uma comparação dos resultados da realização de clusterização através da ferramenta *k-means*, utilizando as distâncias Euclidiana e Correlação de Pearson mostram também mais uma vantagem da padronização dos dados. Os resultados se mostram praticamente iguais, dependendo apenas de um fator constante, seja para a utilização da distância euclidiana quadrada com atributos padronizada, seja para utilização distância de similaridade de coeficiente de correlação de Pearson. Portanto, uma vez que os dados são padronizados, a escolha entre as Distâncias Euclidiana e Correlação de Pearson é simplificada.

Por último, após a padronização dos atributos que serão utilizados como parâmetro de entrada para realizar a clusterização, pode-se dar pesos a cada atributo, isso quer dizer que é possível determinar quais atributos são mais influentes para a formação dos clusters.

Por exemplo, o algoritmo *k-means* assume que cada atributo da amostra contribui com o mesmo grau para o cluster final, porém, dependendo do que se quer analisar, alguns atributos podem desempenhar mais influência. Para estes casos, Xu *et al.* (2018) defendem que a atribuição de diferentes pesos para os atributos pode ajudar na otimização e formação dos clusters. Embora essa etapa seja importante e possa ajudar na etapa de clusterização, ela não é discutida na literatura que utiliza dados de *Smart Card*, sendo encontrada apenas em literaturas que utilizam ferramentas de clusterização em geral.

A literatura atribui essa ponderação de atributos como uma etapa de pré-processamento de dados e como uma etapa que pode fortalecer ou enfraquecer os atributos nas aplicações de técnicas de mineração de dados, como algoritmos de classificação e agrupamento. Ou seja, os atributos mais importantes recebem um peso maior, enquanto os atributos menos importantes recebem um peso menor.

De acordo com Song *et al.* (2007), esses pesos podem ser determinados por dois métodos: com base no conhecimento sobre a importância dos atributos para a análise do fenômeno ou podem ser determinados por algum método estatístico. Por exemplo, Song *et al.* (2007) utilizam o inverso do coeficiente correlação múltiplo como o peso do atributo. Os autores afirmam que a metodologia de pesos escolhida por eles ajudou na melhoria da precisão dos resultados de agrupamento e na eficiência do algoritmo uma vez que foi diminuindo a quantidade de número de interações.

3.4 Conclusões e lacunas

Os resultados encontrados na literatura indicam que as diversas ferramentas de clusterização utilizadas podem ajudar na estimação de padrões de mobilidade a partir de características individuais ou de viagens. Porém, percebe-se que existe uma limitação na literatura, envolvendo a relação entre a aplicação das metodologias e a compreensão do fenômeno, principalmente porque não é discutido de forma satisfatória quais são os impactos causados nos resultados das análises pelas alterações realizadas nos parâmetros de entrada e especificidades de cada ferramenta de clusterização.

A literatura também não discute como esses métodos de clusterização podem contribuir para os problemas de caracterização dos padrões de mobilidade e identificação dos problemas de mobilidade, essas são duas discussões que esta pesquisa procura se aprofundar.

No tocante aos dados disponíveis para Fortaleza (*Smart Card*, GPS, Cadastro de bilhete único e GTFS), assim como os dados utilizados em alguns estudos apresentados nessa revisão, encontra-se grande potencial em obter informações que permitam a utilização de características de viagem e socioeconômicas (endereço de residência) para construir os perfis dos usuários. A identificação de padrões de mobilidade a partir dessas características se mostrou uma alternativa complementar para a etapa de caracterização da mobilidade, que vem se limitando ao longo dos anos à utilização dos dados de pesquisas domiciliares.

Conforme a discussão proposta no capítulo anterior, os estudos que caracterizam a mobilidade sofrem alguma dificuldade de representar a heterogeneidade da demanda, acredita-se que essa dificuldade pode ser superada por uma caracterização da mobilidade a partir da identificação de padrões de mobilidade, ou seja, de grupos de usuários que possuam características de viagens e socioeconômicas em comum.

A revisão realizada permitiu identificar alguns tipos de grupos com relação ao padrão temporal de viagens: i) usuários pendulares, que usam o sistema com propósito de acessar atividades de trabalho ou educação em horários de pico; ii) usuários que apresentam um uso temporal irregular (não usam nos mesmos horários ao longo da semana, mas possivelmente acessam as mesmas atividades), porém com um uso frequente do sistema; e iii) usuários com uso temporal e frequência de utilização irregular (usuários esporádicos). Os tipos de padrões identificados na literatura, em conjunto com os problemas de mobilidade permitirão levantar

hipóteses sobre os padrões de mobilidade em Fortaleza, assim como esses padrões estão distribuídos e mudando ao longo do tempo.

Os atributos que serão utilizados para construir os perfis de usuários e que servirão de parâmetros de entrada para realização da clusterização terão por base principal os objetivos desta pesquisa, os atributos utilizados nesta revisão de literatura, e as características particulares das bases de dados utilizadas. Na definição desses atributos também será considerado os fatores que impactam a mobilidade dos diferentes padrões, fatores que terão como base as hipóteses levantadas. Além disso, a decisão dos pesos de todos os atributos para a definição dos clusters será definida a partir de análises descritivas e análises de sensibilidade, assim como sugerido pelas melhores práticas encontradas na literatura.

A literatura também demonstrou também que ainda são incipientes os estudos que analisam a evolução dos padrões com foco no problema da queda da demanda do sistema de transportes público. Acredita-se ser importante identificar quem são usuários (padrões de mobilidade) que podem estar ingressando ou deixando o sistema ao longo do tempo, como o padrão desses usuários muda ao longo do tempo, assim como identificar onde esses usuários estão localizados no espaço (endereço ou zona de origem da viagem).

A análise temporal e espacial proposta neste estudo tem o foco principal de compreender melhor a queda da demanda. Essas análises buscam responder perguntas específicas que ajudarão a entender melhor o problema de queda da demanda, por exemplo, quais usuários são mais vulneráveis a esta queda da demanda? A partir da identificação de padrões, como esses usuários estão distribuídos espacialmente e temporalmente? Quais os fatores podem estar influenciando cada usuário (padrão de mobilidade)? Essa é uma etapa que pode ser importante para fornecer subsídio para quem tem poder de decisão. Também é intenção desta pesquisa investigar como os problemas de queda da demanda para os diferentes padrões identificados se relacionam com os problemas de mobilidades tipificados por Garcia (2018).

4 CONTEXTUALIZAÇÃO DOS SISTEMAS ANALISADOS

Este capítulo busca expor um conjunto de informações sobre os sistemas (Uso do solo, Transportes e Atividades) que serão analisadas no decorrer desta pesquisa a partir de uma revisão de literatura na caracterização de aspectos da acessibilidade e mobilidade de Fortaleza.

Essa revisão é um esforço resultado de pesquisa do grupo MITUS@PAMUS (Modelagem Integrada dos Transportes e Uso do Solo no Planejamento da Acessibilidade e Mobilidade na Urbe Sustentável), do Programa de Pós-Graduação em Engenharia de Transportes da Universidade Federal do Ceará (PETRAN-UFC). A compreensão desses fenômenos de uso do solo e da oferta do sistema de transportes é essencial para a caracterização dos padrões de mobilidade dos usuários de transporte público por ônibus em Fortaleza, pois elas darão base para a construção de hipóteses sobre os fatores que impactam os padrões.

De acordo com Andrade (2016) e Lima (2017), Fortaleza enfrenta um crescimento acentuado e desordenado que tem empurrado a população de baixa renda para as periferias. Lima (2017) aponta que, a partir de 1980, inicia-se um processo de expulsão da população de classes mais pobres para zonas mais periféricas nas grandes cidades que se chama espraiamento urbano.

O processo de desenvolvimento social e econômico dessas regiões não ocorre na mesma velocidade, ou seja, este processo acaba afastando essa parcela da população das oportunidades de trabalho, educação e outras atividades. O espraiamento urbano descrito por Lima (2017) afasta parte da população das ofertas de empregos e de outras atividades, criando nesses indivíduos a necessidade de realizar mais viagens de longas distância.

Lima *et al.* (2019) constroem um diagnóstico tornando possível perceber a existência de duas cidades que compartilham a mesma centralidade de oferta de empregos: (1) a primeira da população de alta renda, em que esse segmento populacional concentra-se em maior intensidade, acrescida também do lado leste, para onde esse grupo vem consolidando seu vetor de autosegregação e onde os empregos já apareceram em quantidade necessária; (2) a segunda da população de renda baixa, localizada no lado oeste, vetor de periferização cada vez mais consolidado pela segregação residencial imposta à população de baixa renda, especialmente, distante das oportunidades de emprego.

Em parte, esses fenômenos de desigualdades de acesso podem ser considerados resultantes de uma política falha de distribuição e uso do solo e dos dois primeiros paradigmas de

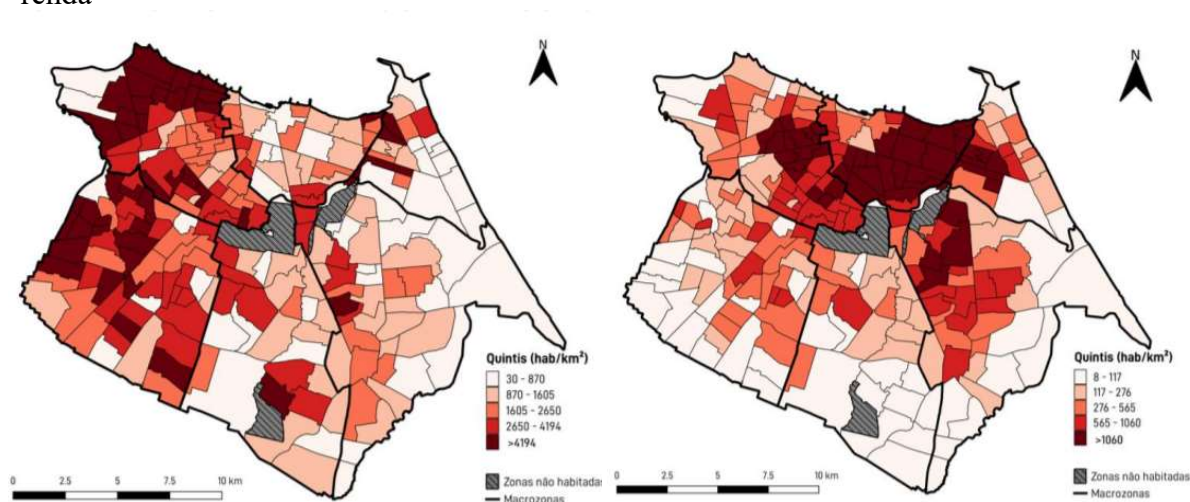
planejamento de transportes que focaram em problemas operacionais de transportes e no aumento da oferta de viagens (SOUSA, 2019).

4.1 Uso do solo e atividades

Como contextualizações do sistema de uso do solo e de atividades serão apresentadas as análises realizadas por Pinto (2020) de distribuição espacial da população e de oportunidade de empregos de Fortaleza com dados de censo de 2010 extrapolados para o ano 2017. O indicador utilizado pelo autor foi densidade de indivíduos por zona (hab/km^2) por faixa de renda. O primeiro mapa da Figura 8 mostra a distribuição para a população classificada como baixa renda, nota-se que as zonas pertencentes ao quintil de maior densidade populacional concentram-se, predominantemente, nas regiões Noroeste e Sudoeste.

Isso reforça a tendência de um processo de periferização desse grupo populacional em direção aos anéis mais periféricos da região noroeste e sul da capital, ao longo dos últimos 20 anos, como constatado pelos trabalhos de Andrade (2016) e Lima (2017). No segundo mapa da Figura 9, Pinto (2020) apresenta a distribuição espacial da população de alta renda, em que é possível visualizar em quais regiões os indivíduos de alta renda concentram-se, majoritariamente, nas zonas mais centrais, percebe-se também que essa densidade reduz à medida que se aproxima das zonas mais periféricas.

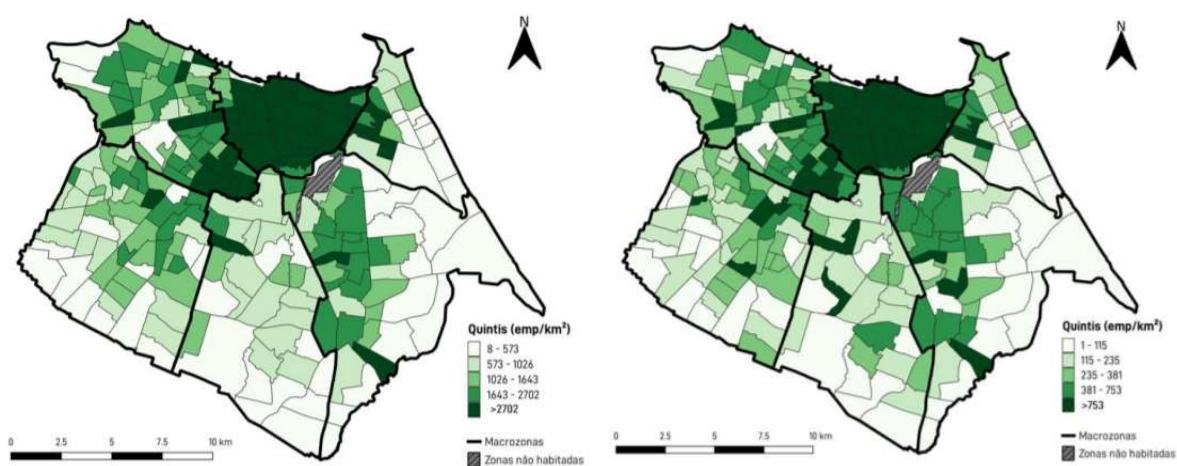
Figura 8 – Mapa de quintis de densidade populacional da população de baixa renda e alta renda



Fonte: Adaptado de Pinto (2020).

Como indicador para caracterizar a distribuição de empregos, Pinto (2020) utilizou a densidade de vagas de emprego (emp/km^2) em relação à população de baixa renda, conforme Figura 9. Como pode ser observado, há uma grande concentração de empregos nas regiões centrais da cidade tanto para a população de baixa como de alta renda. Percebe-se também uma escassez de oportunidades de empregos em parte reforçada da região periférica, em especial, nas zonas a Sudoeste, Sul e Sudeste. Os resultados dessas distribuições permitem observar a existência de uma desigualdade de acesso entre as populações de baixa e alta renda, uma vez que a população de baixa renda reside em zonas periféricas e grande parte das oportunidades de emprego se encontra em zonas mais centrais.

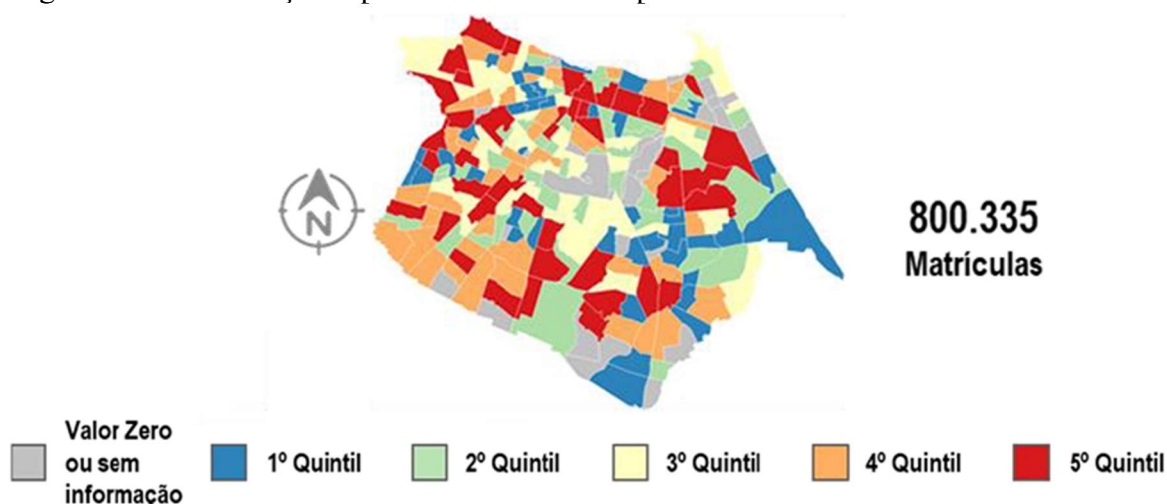
Figura 9 – Mapa de quintis da densidade de empregos da população de baixa renda e alta renda



Fonte: Adaptado de Pinto (2020).

Pinto (2020) não avalia o acesso a oportunidades de educação, para avaliar esses aspectos das oportunidades utilizaremos a análise de Sousa (2019), que utiliza a distribuição das matrículas para o ano de 2015 (Figura 10). Percebe-se uma distribuição espacial das matrículas bem menos concentrada do que a distribuição observada para os empregos analisados por Pinto (2020). De acordo com Sousa (2019), essa distribuição menos concentrada é, fortemente, influenciada pela localização das instituições de ensino básico oriundas do sistema público, que acabam por distribuir-se por quase toda a cidade.

Figura 10 – Distribuição espacial das matrículas para o ano de 2015



Fonte: Adaptado de Sousa (2019).

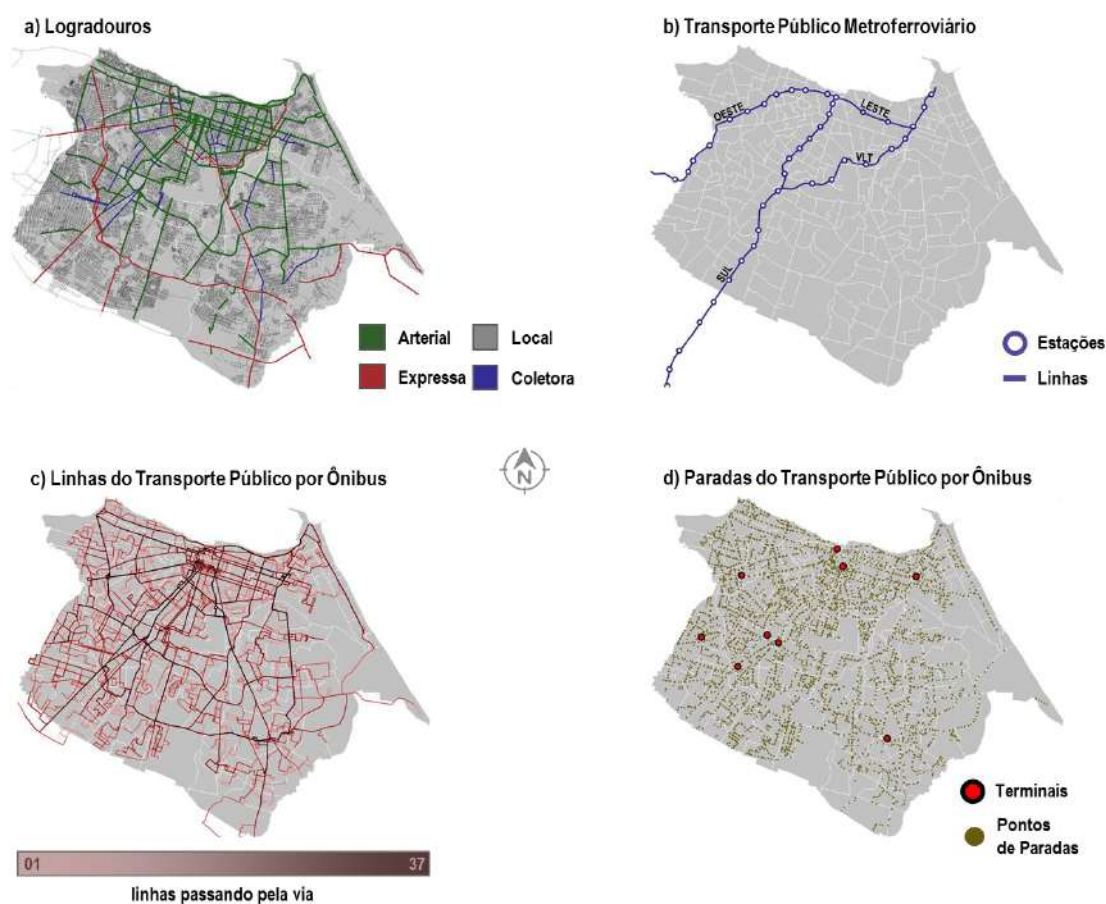
4.2 Sistemas de transportes

Para sintetizar a oferta dos sistemas de transportes serão apresentados mapas de Sousa (2019) que podem ser observados na Figura 11. No primeiro mapa “a)” fica claro uma concentração de vias arteriais na região central da cidade, essas vias que usualmente possuem maior capacidade de deslocamento.

No mapa “b)”, é apresentado o sistema metroviário de transporte coletivo da cidade chamado Metrô de Fortaleza (METROFOR). A Linha Oeste do metrô foi a primeira a ser implantada, operando mais próximo a sua capacidade, porém com *headway* altos que variam de 30 a 90 minutos. A linha Sul foi implantada em 2015, mas com *headway* que varia de 15 a 20 minutos, a linha do VLT foi totalmente concluída em 2020, mas opera parcialmente em fases de teste desde 2018, por fim, a linha Leste foi retomada sua construção apenas em 2019.

Pelos mapas “c)” e “d)”, percebe-se que o sistema de transporte público por ônibus possui uma cobertura de praticamente todo o município de Fortaleza, com uma maior intensidade dessas linhas nas vias que realizam uma conexão entre os terminais.

Figura 11 – Oferta do sistema de transporte

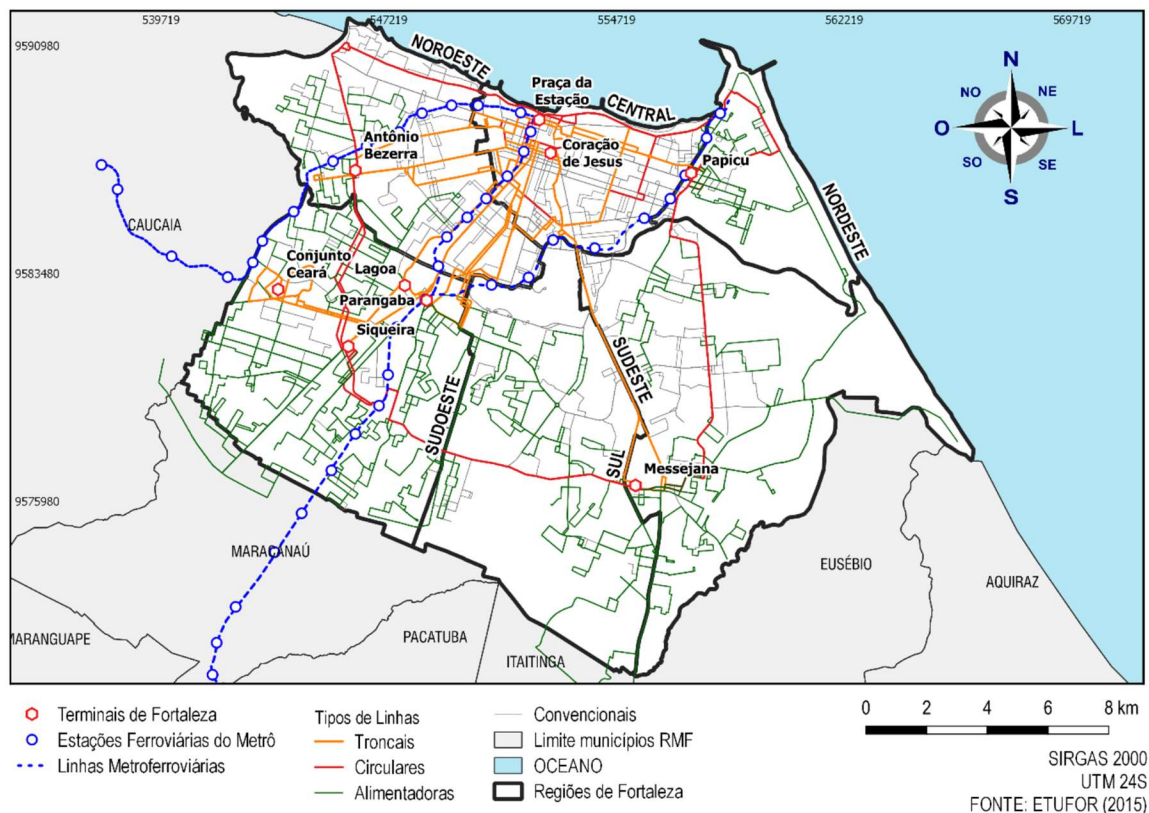


Fonte: Sousa (2019).

4.2.1 Sistema de transporte público por ônibus de Fortaleza

Esse sistema é o objeto de estudo principal desta pesquisa, conhecido como Sistema Integrado de Transporte de Fortaleza (SIT-FOR), é um sistema do tipo tronco-alimentador, composto por 269 linhas troncais e alimentadoras, 19 linhas complementares e 7 terminais de integração física (Fortaleza, 2020). O sistema tronco-alimentador ocasiona muitas as viagens as quais se originam nas zonas periféricas da cidade, não impedindo que muitos usuários ainda passem pelos terminais físicos de integração, o que resulta em tempos maiores de viagem e de espera. Para representar este sistema foi utilizado um mapa de Sousa (2019) que pode ser visualizado na Figura 12. Também consta neste mapa as linhas metroforroviárias.

Figura 12 – Rede de transporte público de Fortaleza



Fonte: Sousa (2019).

A rede possui elevada cobertura (Figura 11c), estando, praticamente, todo o território da cidade a até 500 metros de uma parada de ônibus. Apesar dessa cobertura, Sousa (2019) diz que a conectividade da rede é pequena em grande parte da cidade, uma vez que os terminais são o principal método de integração utilizado. Isso leva a regiões distantes dos terminais (Nordeste, Sudeste e Sul) serem prejudicadas, demonstrando como as linhas troncais privilegiam alguns corredores através de um acesso mais rápido e frequente aos terminais e ao centro da cidade.

O SIT-FOR possui alguns sistemas que são responsáveis por gerar o *Big Data* de Transporte Público de Fortaleza. O sistema de bilhetagem eletrônica, que permite aos usuários a validação do seu embarque, esse sistema é responsável por registrar as viagens de todos os usuários que passam pelas catracas dos veículos.

O sistema aceita 5 formas de pagamento: bilhete único de vale transporte; carteira de estudante; gratuidade de idosos; deficientes; bilhete de funcionários e inteira (pagamento em dinheiro). Existe também o sistema de localização automática de veículos que utiliza

equipamentos de GPS em toda a frota de transporte público de Fortaleza, este registra a localização de cada veículo do transporte em períodos temporais de 30 segundos.

Por fornecerem informações com alto grau de precisão e em tempo real, esses dois sistemas apresentam grandes vantagens para os usuários e para os gestores. Por causa desses sistemas, os usuários têm um processo de embarque e desembarque mais ágil, além da possibilidade de integração fora do terminal. Para os gestores, com essas informações é possível identificar e agir para problemas em tempo real, assim como utilizar esses dados para o planejamento futuro da oferta.

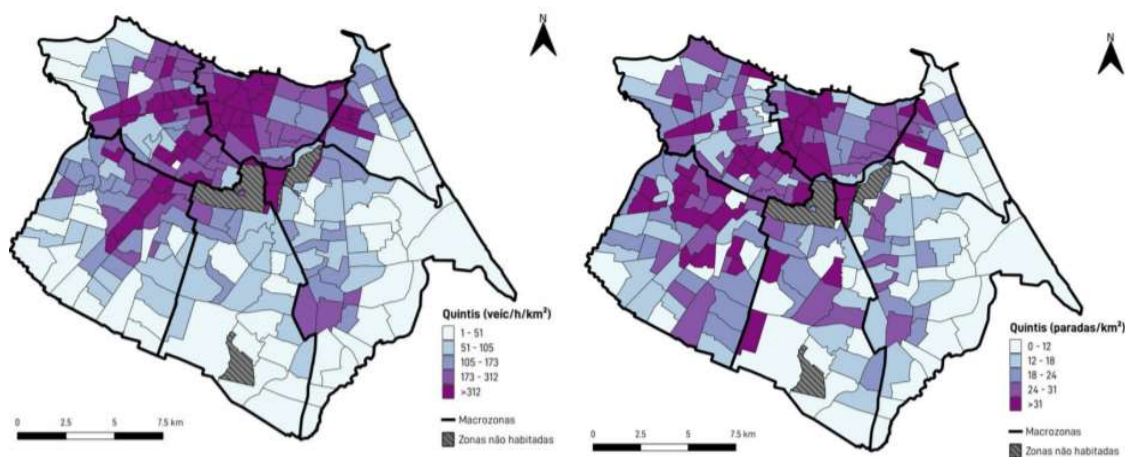
Sousa (2019) também traz algumas intervenções que aconteceram entre os anos de 2014 e 2019 as quais tiveram como objetivo garantir a esta rede de transporte público melhores níveis de qualidade e acesso: a) implantação de linhas que operam com veículos biarticulados, em um sistema BRT, funcionando restritamente às conexões do terminal do Papicu e da região Central ao terminal do Antônio Bezerra; b) ampliação da quantidade de faixas exclusivas de ônibus, capazes de aumentar as velocidade médias de deslocamentos nos principais corredores arteriais da cidade; c) implantação de sistema de bicicleta compartilhado, que se apresenta como mais uma alternativa para o acesso e para a difusão do transporte público.

Para complementar as características de oferta do sistema de transporte público por ônibus serão apresentados dois indicadores analisados por Pinto (2020): Frequência de linhas transporte público por zonas e Densidade de pontos de parada por zona (Figura 13). O primeiro indicador traduz a característica de frequência da rede e foi definido como a soma do fluxo horário, no período de pico, das linhas de transporte público por km² da zona. Pinto (2020) sugere que quanto maior o valor desse indicador melhor serão os níveis de acessibilidade da zona, dado que ele reflete no tempo de espera pelo veículo.

A distribuição visualizada no primeiro mapa da Figura 13 era esperada por Pinto (2020), dado que uma população de baixa renda representa a principal demanda por esse serviço e são nessas regiões em que esse grupo socioeconômico está mais adensado. Também é possível notar uma baixa frequência nas áreas mais extremas da cidade, o que é reconhecido por Pinto (2020) como uma condição de agravamento da acessibilidade, em especial, para os usuários situados na parte mais externa da região sudoeste, devido essa ser uma área de concentração de pessoas de baixa renda e se apresenta como um dos principais polos de construção de habitações populares pelo Poder Público.

O segundo indicador (densidade de pontos de parada por zona) da Figura 13 se propõe a caracterizar a cobertura da rede de transporte público. Pelo mapa de quintis, é possível perceber que em termos de cobertura, há uma melhor distribuição da rede de transporte público. Apesar de ainda ser possível distinguir uma melhor cobertura nas zonas central e oeste, não se verifica uma desigualdade tão grande com relação às regiões mais periféricas dessas regiões. Dessa forma, essa análise complementa a análise de Sousa (2019), e pode-se dizer, em termos de cobertura a rede de transporte público, que esse sistema atende de forma satisfatória as zonas mais adensadas de indivíduos de baixa renda.

Figura 13 – Mapa de Quintis da densidade de frequência de linhas e pontos de parada



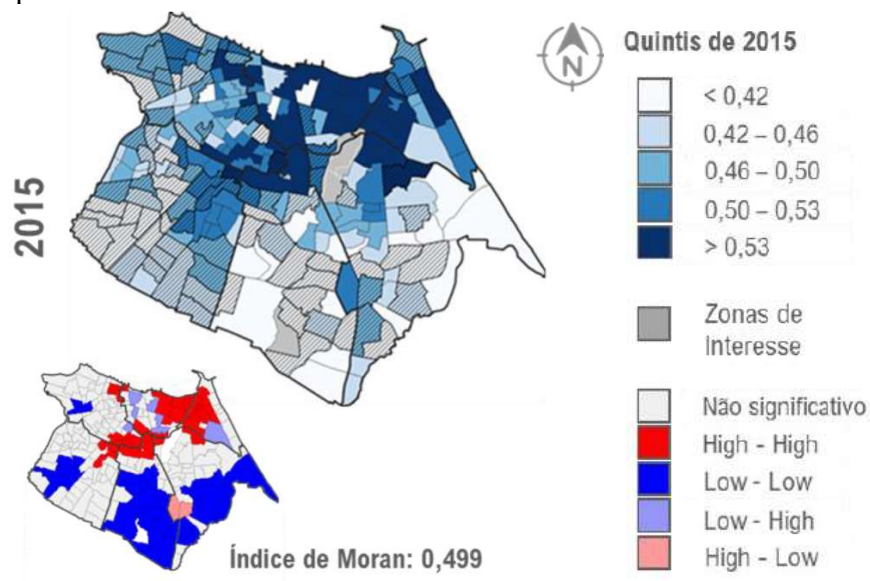
Fonte: Adaptado de Pinto (2020).

4.3 Problemas de mobilidade para a população de baixa renda

Sousa (2019), para avaliar a acessibilidade de Fortaleza, utiliza um indicador de acessibilidade que considera o nível de serviço, quantidade de oportunidades de empregos e de domicílios de toda a região, momento temporal de interesse e os grupos sociais. O autor realiza uma análise da distribuição espacial deste indicador. Para 2015, a distribuição apresentou uma concentração das zonas com os melhores níveis de acessibilidade dentro ou próximas à região Central, sendo essa a região com maior oportunidades de emprego. Também é possível observar uma maior concentração de zonas com acessibilidade intermediária na região Sudoeste do que na região Sudeste, sendo a primeira a área da cidade servida por diversos terminais de transporte coletivo. O autor destaca ainda um distanciamento da população de baixa renda, que residem mais

nas zonas periféricas (destacadas como zonas de interesse no mapa da Figura 14), das zonas com melhor acessibilidade.

Figura 14 – Distribuição espacial da acessibilidade da baixa renda para 2015



Fonte: Adaptado de Sousa (2019).

Conforme discutido na revisão de literatura, a mobilidade pode ser impactada pela acessibilidade. Assim, baixos níveis de acessibilidade podem indicar baixos níveis de mobilidade. Ainda que não seja objetivo desse estudo avaliar a relação de causalidade entre esses dois aspectos, entende-se que a população de baixa renda, que reside, principalmente, em zona periféricas, e que apresentam baixos níveis de acessibilidade, pertencem a um grupo mais vulnerável da população. Consequentemente, a mobilidade dos usuários que residem nessas regiões é mais vulnerável a as alterações que acontecem nos sistemas de transportes, uso do solo e atividades.

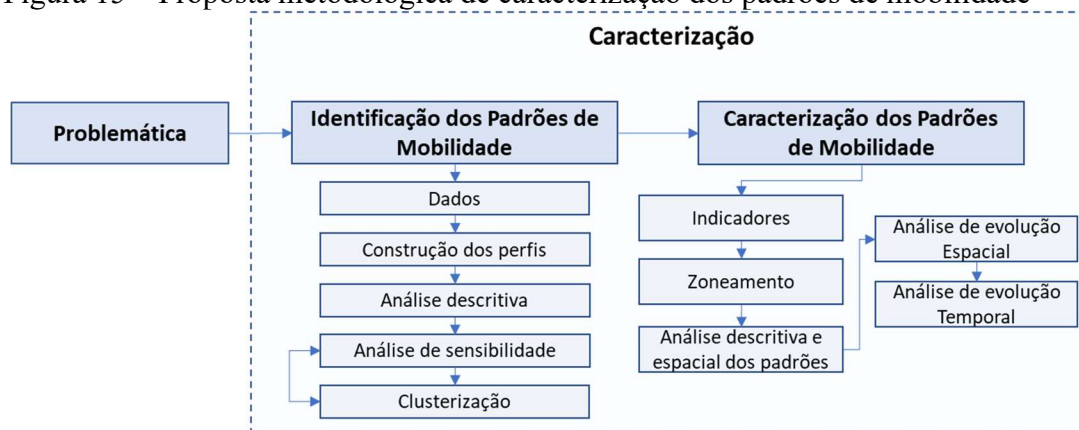
No geral, os estudos realizados pelo grupo de pesquisa PAMUS@MITUS buscam caracterizar e diagnosticar os problemas de acessibilidade e mobilidade da população de baixa renda. Esses estudos evidenciam uma clara desigualdade nas condições de acessibilidade e mobilidade para essa população. Por tanto, seguindo esta mesma linha, é de interesse deste estudo avaliar problemas de mobilidades que afetam a população de baixa renda.

5 PROPOSTA METODOLÓGICA PARA IDENTIFICAÇÃO E CARACTERIZAÇÃO DOS PADRÕES DE MOBILIDADE A PARTIR DE DADOS DE *SMART CARD*

No capítulo 2, foi realizada uma revisão de literatura cujo objetivo era identificar lacunas no processo de caracterização da mobilidade. No capítulo 3, é feita uma revisão de literatura sobre o uso de dados de *Smart Card*, GPS e GTFS no planejamento de transportes com foco no planejamento de transporte público por ônibus. Ainda no capítulo 3, também é realizada uma revisão de como a identificação de padrões utilizando essas novas fontes de dados vem sendo utilizada no planejamento de transportes e como podem ajudar nas lacunas identificadas no capítulo 2. No capítulo 4, realizamos uma breve revisão sobre a caracterização dos principais sistemas, que serão analisados nesta pesquisa. O propósito dessa revisão do quarto capítulo é reunir informações que darão base para propor e analisar as hipóteses de pesquisa. As lacunas identificadas e como esta pesquisa pretende preenchê-las serão mais detalhadas nas descrições das etapas deste método.

Nesse sentido, o método é composto de quatro etapas principais, sendo elas: a primeira é a identificação da problemática, nessa etapa se propõe representar, visualmente, a problemática que será investigada; a segunda etapa tem como objetivo detalhar os dados que serão utilizados; a terceira etapa tem como objetivo identificar os padrões de mobilidade evidenciando a heterogeneidade da demanda; a última etapa é de caracterização dos padrões de mobilidade, em que serão propostos indicadores e análises para relacionar os problemas de mobilidade com os padrões identificados. A cronologia dessa metodologia pode ser verificada na Figura 14.

Figura 15 – Proposta metodológica de caracterização dos padrões de mobilidade



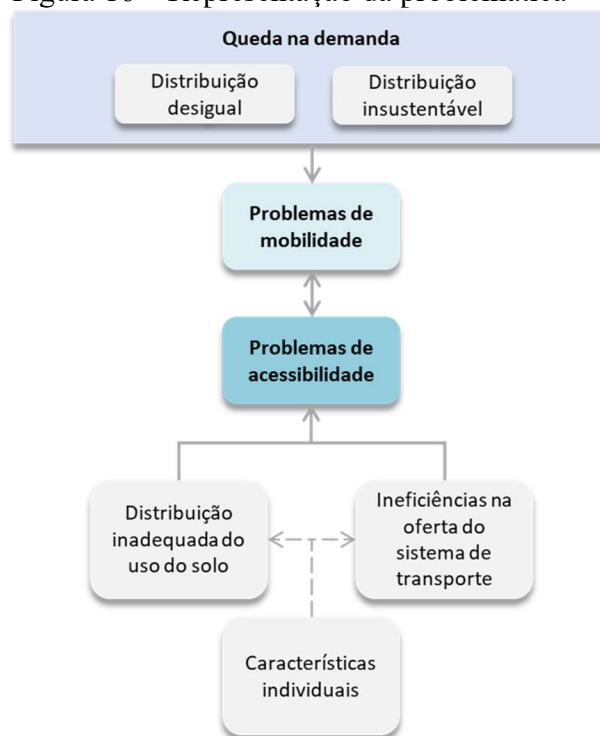
Fonte: Elaborada pelo autor.

5.1 Problemática

Esta pesquisa se propõe a investigar a problemática de queda na demanda de transporte público por ônibus em Fortaleza. Conforme apresentado na literatura, problemas de queda na demanda são naturalmente associados a problemas de mobilidade (GARCIA, 2016). A partir dos problemas tipificados por Garcia *et al.* (2018), para a problemática de interesse, é possível investigar problemas de distribuição desigual e insustentável da mobilidade.

Como representação da problemática (Figura 16) propõe-se uma adaptação das hipóteses de causalidade entre mobilidade e acessibilidade sugeridas por Garcia *et al.* (2018). Portanto, a problemática de interesse é a compreensão de como a queda na demanda do transporte público por ônibus em Fortaleza se relaciona com os problemas de mobilidade tipificados por Garcia *et al.* (2018). Para isso, este método propõe a caracterização de diferentes padrões de mobilidade como modo de considerar a heterogeneidade da demanda de uma mesma zona. Esses padrões serão identificados a partir de um método de clusterização que utiliza características da mobilidade desses usuários para agrupá-los.

Figura 16 – Representação da problemática



Fonte: Adaptado pelo autor a partir de Garcia *et al.* (2018).

A partir da revisão de literatura realizada, acredita-se que existe uma heterogeneidade na demanda por transportes. Por tanto, a investigação desses problemas terá como base uma hipótese central a existência dessa heterogeneidade em relação a intensidade e a regularidade de uso do sistema de transporte público por ônibus, ou seja, diferentes níveis de mobilidade podem ser observados dentro de uma mesma zona residencial ou de um mesmo grupo social. Nesse estudo, a heterogeneidade da demanda será representada por padrões de mobilidade. A partir dessa hipótese central, um conjunto de questionamentos é levantado.

Em Fortaleza, grande parte dos usuários cativos do sistema são usuário de classe baixa e média e a maior parte dessa população reside em zonas periféricas, conforme apontado por Henrique (2004) e por Sousa (2019). Por tanto, espera-se que usuários de padrões de mobilidade que apresentam características de uso regular e dependente do sistema estão mais presentes nas zonas mais periféricas, e que os usuários de padrões de mobilidade menos regulares e menos dependentes residem em zonas mais centrais. Conseqüentemente, essas relações implicam em uma distribuição desigual da mobilidade para estes usuários. Por tanto, espera-se que as análises realizadas confirmem as análises realizadas por Henrique (2004).

Por fim, espera-se que os usuários de zonas mais periféricas, por conta dos baixos níveis de acessibilidade dessas regiões, e por fazerem parte de um contexto histórico e socioeconômico que tende a penalizar mais severamente a população de baixa renda, sejam mais vulneráveis a uma redução da mobilidade. Assim, espera-se que a queda na demanda seja mais intensa nas regiões periféricas. Um resumo das hipóteses está listado na tabela 2. Elas irão guiar as análises propostas neste método e podem ser classificadas em dois tipos.

Tabela 2 – Lista de hipóteses sobre os padrões de mobilidade

Hipóteses de Pesquisa	
Padrões de mobilidade esperados	1) Diferentes padrões de mobilidade são esperados para os usuários de transporte público por ônibus em Fortaleza.
Relações entre os padrões de mobilidade e os problemas de mobilidade	2) Os padrões de mobilidade que apresentam características de uso dependente e regular estão proporcionalmente mais presentes em zonas mais periféricas e usuários pertencentes a padrões de uso não dependente e menos regular residem em zonas mais centrais. 3) A queda na demanda ocorre com mais intensidade em padrões de mobilidade de uso dependente e mais regular, e com menor intensidade para padrões não dependentes e menos regulares

Fonte: Elaborado pelo autor.

5.2 Dados

A segunda etapa é a organização e tratamento dos dados utilizados, este processo é essencial para viabilizar a construção de um conjunto de dados que permita realizar as análises que serão propostas no decorrer dessa metodologia. Conforme apresentado na revisão de literatura, serão utilizadas quatro bases de dados que formam o *Big Data* de Transporte Público de Fortaleza: Dados de *Smart Card*; Dados de GPS; Dados de GTFS e Dados de Cadastro do Bilhete Único. A seguir serão apresentados os tratamentos realizados em cada uma das bases.

5.2.1 Dados de *Smart Card* e GPS

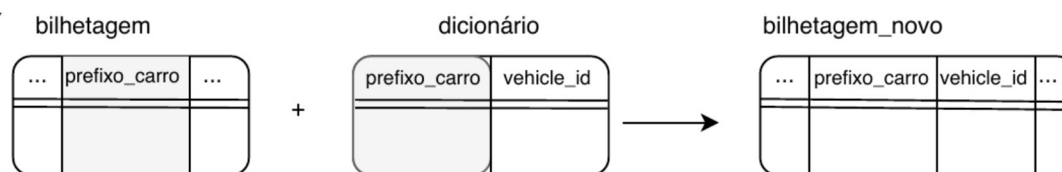
Os dados de bilhetagem contêm 9 colunas, sendo originalmente divididos em diferentes arquivos, em que cada arquivo armazena um dia inteiro de validações. As colunas são:

- *id*: representa o código de identificação do cartão e do usuário;
- *linha*: representa o código da linha a ser utilizada durante a validação;
- *nome_linha*: nome da linha utilizada;
- *prefixo_carro*: nome do veículo;
- *hora*: hora e dia da validação;
- *tipo_cartao*: código do tipo de cartão utilizado;
- *nome_cartao*: nome do tipo de cartão utilizado;
- *sentido_viagem*: o sentido da viagem (ida ou volta);
- *integração*: se a viagem foi integração ou não.

A coluna *tipo_cartao* contém 9 tipos diferentes, para facilitar as análises futuras esses tipos serão agrupados em quatro tipos principais: vale transporte, carteira de estudante, gratuidade e inteira; as colunas *nome_linha* e *nome_cartao* são removidas por conterem informações correspondentes nas colunas *linha* e *tipo_cartao*; os dados de bilhetagem não trazem nenhuma informação espacial, sendo necessário integrá-los com os dados de GPS. Para isso, foi aplicada a metodologia proposta por Braga (2019) de integração e georreferenciamento dos dados de bilhetagem, a partir dos dados de GPS, essa metodologia é responsável por estimar a localização das validações através do horário de validação.

Para integrar as bases de bilhetagem e GPS, Braga (2019) utiliza um dicionário disponibilizado pelos órgãos operadores do sistema. A base da bilhetagem contém a coluna `prefixo_carro`, enquanto a base do GPS possui a coluna `vehicleid`, ambas representando a identificação do veículo em questão. Esse léxico apresentado no dicionário possui uma “tradução” das duas colunas, com um `vehicleid` para cada `prefixo_carro`, faz-se uma junção da bilhetagem com esse dicionário para criar a coluna `vehicleid` na base da bilhetagem (Figura 16).

Figura 16 – Integração das bases de bilhetagem e GPS



Fonte: Braga (2019).

Por fim, uma vez que as duas bases são relacionáveis através da coluna `vehicleid` igual ao do GPS, Braga (2019) realiza a integração. Então, para o mesmo `vehicleid`, é buscado o momento mais próximo da bilhetagem na base do GPS, a partir desse encontro, a coordenada aproximada do ônibus é extraída da base de GPS e integrada na base da bilhetagem. Na Tabela 3, é possível verificar uma amostra desses dados de bilhetagem já relacionados com os dados de GPS.

Tabela 3 – Amostra dados de bilhetagem

id	linha	prefixo_carrão	hora	sentido_viagem	tipo_cartao	integracao	lon	lat
5298509	61	36001	2018-10-01T06:44:54Z	Volta	Vale transporte	N	-38.5381	-3.7226
4611338	61	36001	2018-10-01T06:53:23Z	Volta	Estudante	S	-38.534	3.72792
6854070	61	36001	2018-10-01T06:55:45Z	Volta	Vale transporte	N	38.5322	3.72862
2128321	61	36001	2018-10-01T06:55:49Z	Volta	Vale transporte	N	38.5322	3.72862
7361166	61	36001	2018-10-01T06:55:57Z	Volta	Vale transporte	N	38.5321	3.72868
6596413	61	36001	2018-10-01T07:06:56Z	Volta	Vale transporte	N	38.5224	3.72918
4904265	61	36001	2018-10-01T07:08:56Z	Volta	Vale transporte	N	38.5225	3.72926
3530328	61	36001	2018-10-01T07:09:27Z	Volta	Gratuidade	N	38.5225	3.72925
6596413	61	36001	2018-10-01T07:24:57Z	Volta	Vale transporte	N	38.5413	3.72118
3666176	61	36001	2018-10-01T07:26:21Z	Volta	Vale transporte	N	-38.537	3.72295

Fonte: Elaborado pelo autor.

Serão utilizados os dados de validações para os anos de 2014 e 2018. O ano de 2014 foi escolhido porque o bilhete único começou a ser implantado no ano de 2013, fazendo que a população aderisse a utilização do sistema somente no decorrer do ano desse mesmo ano, ou seja, o retrato dos dados para 2013 é incompleto.

Em relação a 2018, justifica-se nossa escolha por ser o último ano disponível com informações completas dos dados pela Prefeitura de Fortaleza. Além disso, um conjunto de políticas de transportes foi realizado durante o período e mesmo com essas políticas a demanda para esse período apresenta queda relevante. O que torna este período de 5 anos (01/2014 – 12/2018) interessante para a investigação da problemática de interesse.

5.2.2 Dados de GTFS

A Especificação Geral de Feeds de Transporte Público (GTFS) especifica um formato padrão para compartilhamento de dados de transporte público pelas agências reguladoras. Esses dados são fornecidos para o Google e contêm informações de horários, itinerários das linhas e tarifas, permitindo que a empresa faça uso desses dados nos seus produtos (como no Google Maps)

para fornecer rotas, estimações de tempo e custo monetário para o usuário. Esses dados não precisam passar por nenhum tipo de tratamento. Os dados de GTFS são divididos em cinco principais arquivos:

- *stops.txt*: as paradas do sistema de transporte público e sua localização;
- *routes.txt*: informações básicas de cada linha de transporte público;
- *trips.txt*: as viagens realizadas por cada linha;
- *stop_times.txt*: o horário programado de chegada de cada veículo, em cada viagem, às paradas do sistema;
- *shapes.txt*: os pontos que definem cada rota das linhas do sistema de transporte público.

Esses dados possuem ricas informações sobre a oferta do sistema de transporte, porém, para esta pesquisa só serão utilizados para validar a utilização do endereço de cadastro do bilhete único como ponto de origem das viagens realizadas pelos usuários. Portanto, a partir dos dados de GTFS serão identificadas as paradas pertencentes à linha mais utilizada pelo usuário, que esteja mais próxima da residência dos usuários de cadastro do bilhete único. Essa etapa requer um alto tempo de processamento computacional, assim, será analisado somente o período de um mês dessas validações.

5.2.3 Dados de cadastro do bilhete único

O Bilhete Único Fortaleza é o atual modelo de integração utilizado no transporte coletivo de Fortaleza, com ele, é possível pegar quantos ônibus quiser, no período de duas horas, em qualquer sentido, pagando apenas uma passagem ou meia passagem, no caso dos estudantes. O Bilhete Único é um cartão de crédito pré-pago para o transporte coletivo. O usuário pode inserir o valor que desejar e, ao embarcar em um ônibus ou van, basta passar o cartão no validador que, automaticamente, a passagem será debitada e a catraca será liberada. Com a utilização do Bilhete Único, se o usuário pegar outro ônibus no intervalo de duas horas ele não paga outra passagem (Prefeitura de Fortaleza, 2020). Para realizar a solicitação do Bilhete Único é preciso ser residente de Fortaleza e apresentar os seguintes documentos: comprovante de endereço, cópia e original de identidade, cadastro de pessoa física (CPF).

Essas informações, de cada usuário, são cadastradas e atreladas a um código id, o mesmo utilizado nos dados de bilhetagem, tornando então os dados de cadastro relacionáveis com os dados de *Smart Card*. Para proteger os dados dos usuários, nenhum nome ou documento de identificação (CPF) será explorado durante as análises a serem realizadas, portanto esses dados foram removidos.

Os dados de cadastro do bilhete único são dados acumulativos ao longo dos meses e anos, o que gera vários registros para um mesmo usuário. Assim, foram utilizados os dados acumulados até dezembro 2014 e até dezembro de 2018 para gerar duas bases diferentes, uma para 2014 e outra para 2018. Além disso, para garantir que usuários não possuam mais do que um registro, foram considerados somente o registro mais recente de usuário. Também foram eliminados alguns registros de usuários que possuíam problemas no campo do endereço.

Uma vez que a base inicial de cadastro foi tratada, a última etapa é georreferenciar os endereços identificados para as bases de 2014 e 2018. Para isso, foi utilizada a ferramenta de Geocoding API do Google Maps através da linguagem de programação R no software R Studio. Na Tabela 4, é possível verificar uma amostra desses dados de cadastros já com endereço georreferenciados.

Tabela 4 – Amostra dados de cadastro bilhete único.

id	DataCadastro	Lat	Lon
100002	05/23/18	-3.7831	-38.5101
100013	05/06/18	-3.7265	-38.5669
100033	04/27/17	-3.7744	-38.5576
1000473	06/17/18	-3.7866	-38.5609
100072	04/21/17	-3.7224	-38.5546
100077	06/05/18	-3.7183	-38.5828
100092	05/22/18	-3.7573	-38.5109
100097	04/13/18	-3.7618	-38.5845
100099	05/23/18	-3.8166	-38.5976

Fonte: Elaborado pelo autor.

Para o ano de 2014, foram georreferenciados os endereços de **335.526** usuários, de um total de **1.011.832** usuários rastreáveis que utilizaram o sistema, representando **33,16%** desses usuários. Para o ano de 2018, foi possível georreferenciar o endereço de **363.745** usuários, de um total de **873.553** usuários possíveis de rastreamento que utilizaram o sistema, equivalendo a **41,6%** desses usuários. Os usuários rastreáveis são aqueles que utilizam um cartão que possua um código

de identificação rastreável (Vale Transporte, Estudante, Gratuito). Em 2014, as validações rastreáveis representam 70% de todas as validações, para o ano de 2018, essas validações representam 75% das validações total. O resto das validações é do tipo inteira, que são realizadas sem nenhum tipo de cartão, e, portanto, não podem ser rastreáveis.

Conforme visto na literatura, Freitas (2014) demonstrou que o local de validações não coincide com o endereço de cadastros, portanto, o georreferenciamento dos endereços da base de cadastro busca estimar a origem das viagens realizadas pelos os usuários. Para validar a utilização desses endereços como origem das viagens propõe-se avaliar as seguintes distribuições: distância da parada mais próxima, que faça parte da linha frequente do usuário para a residência dos usuários; distância do cluster das primeiras validações para a residência dos usuários.

Para calcular essas distâncias será utilizado o seguinte método: 1) Selecionar os usuários que realizaram mais do que 20 validações no período analisado (outubro/2018); 2) Identificar o cluster principal de validação desses usuários através da ferramenta DBSCAN; 3) Identificar a parada, que faça parte da linha frequente do usuário, mais próxima ao endereço de cadastro de cada usuário através dos dados de GTFS; 4) Calcular as distâncias entre as paradas identificadas e as residências; 5) Calcular a distância entre o centroide do cluster de validação e as residências dos usuários.

5.3 Identificação dos padrões de mobilidade

Conforme apresentado na literatura, há dois métodos de clusterização que são bastante utilizados para identificar padrões e grupos de usuários através de dados de *Smart Card*: *k-means* e HAC (Clusterização Hierárquica). O HAC agrupamento hierárquico requer, como parâmetro de entrada, uma matriz de distância $N \times N$, sendo N a quantidade de observações a serem agrupadas, ou seja, são necessários o cálculo e o armazenamento dessa matriz de entrada, o que torna esse método inadequado para conjuntos de dados muito grandes. No *k-means*, o cálculo das distâncias de similaridade ocorre durante o seu processo iterativo, tornando esse método muito mais rápido e eficiente do ponto de vista computacional.

Como veremos nas próximas seções, esse método considerará diferentes aspectos para definir a quantidade de grupos esperados, porém uma vez definido, será necessário utilizar uma ferramenta de clusterização que permita a escolha prévia dessa quantidade de grupos.

Desse modo, como serão utilizadas base de dados que possuem em média de um milhão de usuários e, uma vez que o método é amplamente utilizado na literatura, a escolha pelo *k-means* como método de clusterização se justifica por viabilizar os cálculos computacionais necessários e a definição prévia do número de grupos. Além disso, Pieroni (2018) demonstrou que é possível obter resultados semelhantes mesmo utilizando diferentes ferramentas de clusterização.

Portanto, nessa proposta metodológica será utilizar a ferramenta de clusterização *k-means* para identificar padrões de mobilidade a partir de características individuais e de uso do sistema. O *k-means* requer três variáveis de entrada: base de dados com observações e atributos, número de grupos e medida de similaridade. Assim, o primeiro objetivo dessa etapa é definir quais as características dos usuários podem ser obtidas através dos dados disponíveis que podem ajudar a representar a mobilidade desses usuários. A etapa de cálculo e atribuição desses atributos para cada usuário se chama construção de perfis dos usuários, esses perfis serão utilizados na etapa de clusterização para agrupar usuários que possuam características em comum e separar usuários com características distintas.

O objetivo seguinte é definir a quantidade de grupos esperados como resultado da clusterização, essa decisão será resultado de um conjunto de análises que terá como base uma análise descritiva dos perfis dos usuários, os resultados das ferramentas estatísticas de definição da quantidade de grupos ideal apresentadas na literatura e, por fim, as hipóteses sobre os padrões de mobilidade existentes a partir da revisão de literatura. O último parâmetro necessário é a medida de similaridade, nesta pesquisa será utilizada a distância do tipo Euclidiana, que é a mais utilizada pela literatura.

A última etapa antes da clusterização e identificação dos padrões de mobilidade é análise de sensibilidade dos pesos dos atributos, o objetivo dessa análise é definir os pesos que cada atributo terá na formação dos clusters. O método utilizado será de alteração dos pesos de modo arbitrário buscando reduzir os impactos de atributos que possuem alto nível de correlação com outros atributos e aperfeiçoar a distribuição dos atributos para cada um dos grupos. Essa otimização da distribuição dos atributos busca facilitar a interpretação dos grupos, fazendo que as distribuições de frequência dos atributos para cada grupo sejam visualmente distinguíveis e fáceis de interpretação.

O *k-means*, por ser um método não determinístico, pode resultar em diferentes resultados a cada execução. Como a proposta desta pesquisa é analisar os padrões para os anos de

2014 e 2018, foi necessário juntar as bases dos usuários e seus atributos de 2014 e 2018 a fim de obter um único resultado de identificação dos padrões de mobilidade. Após executar o *k-means*, essas bases foram novamente separadas, desse modo, foi possível que a forma dos padrões para os dois anos seja igual e comparável.

5.3.1 Construção de perfis dos usuários

Serão construídos perfis dos usuários para os anos de 2014 e 2018 de forma separada, permitindo, assim, para as análises propostas neste método, a comparação entre os dois anos. Com intuito de representar características de intensidade e regularidade de uso do sistema de transporte público por ônibus são propostas a utilização de três atributos: Frequência média diária de uso, Desvio padrão da frequência média diária de uso, Número de dias de uso durante o ano.

A fim de representar os tipos de atividades dos usuários, assim como a regularidade temporal dessas atividades, propõe-se a utilização de dois atributos: Distância temporal média entre embarques em um dia e Desvio padrão da distância temporal entre embarque. Esses atributos serão calculados a partir das observações dos dados de *Smart Card*, cada uma delas serão detalhadas a seguir:

- **número de dias de uso durante o ano:** com esse atributo é possível captar tanto características de intensidade de uso ao longo do ano, como também, quando associadas com as variáveis anteriores, a regularidade de uso durante o ano (MA *et al.*, 2013). Essa variável foi calculada a partir da contagem do número de dias em que o usuário utilizou o sistema, independente das frequências e excluindo os finais de semana;
- **frequência média diária de uso:** esse atributo representa a intensidade que os usuários usam o transporte coletivo em um dia útil (AGARD E TREPANIER, 2013). Esse atributo permite segregar usuários apresentam características de uso diferentes com relação à frequência e intensidade de uso durante um dia. O cálculo foi realizado por meio da contagem do número de validações diárias realizadas pelos usuários e então realizado a média para os dias úteis. A equação 1 soma todas as validações realizadas nos dias úteis ($\sum_D Validações_{D,i}$) e divide pelo número de dias úteis ($NDias_i$);

$$A_FREQMED_i = \frac{\sum_D Validacões_{D,i}}{NDias_i} \quad 1$$

- **desvio padrão da frequência média diária de uso:** esse atributo representa a variabilidade da frequência dos usuários, ou seja, o objetivo segregar usuários a partir da regularidade de uso diário do sistema. A variável é obtida calculando-se o desvio padrão das observações de frequência diárias. Esse atributo também é utilizado por Pieroni (2018). A equação 2 calcula o desvio padrão populacional da frequência média diária;

$$A_FREQMED_SD_i = \sqrt{\frac{\sum_D (Validacões_{D,i} - \overline{Validacões_i})^2}{NDias_i - 1}} \quad 2$$

- **distância temporal média entre embarques (horas):** esse atributo é calculado como sendo o intervalo temporal entre a primeira e a última validação. A ideia é que através dessa variável seja possível perceber uma estimativa do tempo da atividade dos usuários, assim como feito por Pieroni (2018). A equação 3 calcula a média da diferença de horas entre primeira ($HPValidação$) e a última ($HUValidação$) validação para os dias úteis ($NDias_i$) que foram utilizados;

$$A_DIST_TEMP_MED_i = \frac{\sum_D (HUValidação - HPVValidação)_{D,i}}{NDias_i} \quad 3$$

- **desvio padrão da distância temporal média entre embarques (horas):** esse atributo representa a variabilidade das distâncias temporais, ou seja, o objetivo é segregar usuários que possuam atividades que não tenha um padrão temporal regular ao longo do ano. A variável é obtida calculando-se o desvio padrão das observações das distâncias temporais ao longo do ano. A equação 4 calcula o desvio padrão populacional da distância temporal média.

$$A_DIST_TEMP_SD_i = \sqrt{\frac{\sum_D (DIST_TEMP_{D,i} - \overline{DIST_TEMP}_i)^2}{NDias_i - 1}} \quad 4$$

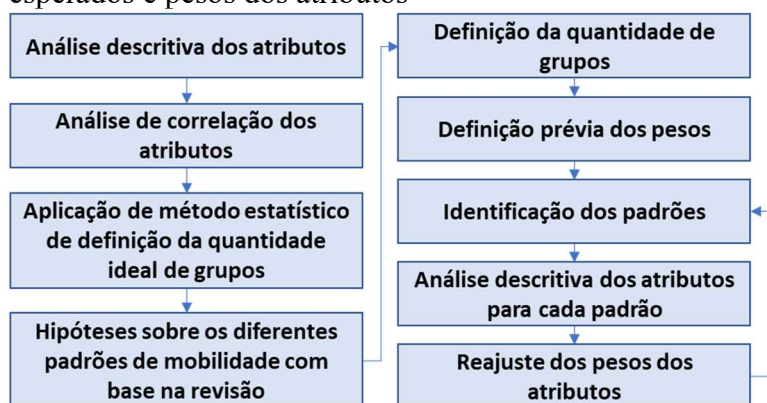
Após calculados os atributos dos usuários foram identificados alguns *outliers*, como usuários que possuíam uma frequência média por dia maior do que seis. Estes usuários representavam um valor menor do que 0.3% dos usuários e foram removidos das análises. Foi considerado que esses usuários representavam alguma inconsistência, como a utilização de um mesmo cartão por mais de uma pessoa no mesmo dia. Foram removidos também usuários tinham o atributo de número de dias durante o ano menor do que 7. Estes usuários representavam um valor menor do que 0.5% das viagens e foram removidos das análises. Para esta análise, foi considerado que esses indivíduos não são, de fato, usuários do sistema durante o período, e que utilizam de forma muito esporádica, podendo atrapalhar a formação dos padrões de mobilidade que importam para a análise.

Os atributos escolhidos possuem unidades, variações e escalas diferentes, e, uma vez que a ferramenta de clusterização *k-means* não consegue perceber essas diferenças, se faz necessário que esses atributos sejam, de algum modo, comparáveis. A literatura traz como solução para essa questão o processo de padronização desses atributos, esse processo remove a média dos atributos escalando estes para uma variação unitária. A padronização será realizada somente antes da utilização desses atributos no processo de clusterização, portanto, as análises descritivas dos atributos serão realizadas antes deste processo.

5.3.2 Definição da quantidade de grupos e pesos dos atributos

O objetivo desta etapa é definir o número de padrões a serem identificados e definir os melhores pesos de cada atributo de modo a ponderar a importância de cada atributo na identificação dos grupos. A definição desses parâmetros é um processo de oito etapas (Figura 17), em que as três últimas são um processo cíclico até encontrar os valores definitivos para os pesos, são elas: análise descritiva dos atributos, análise de correlação entre os atributos, aplicação de método estatístico para definir quantidade de grupos, hipóteses sobre os diferentes padrões, definição da quantidade de grupo com base na análises, definição prévia dos pesos, identificação dos padrões, análise descritiva e reajuste dos pesos.

Figura 17 – Método de definição do número de padrões esperados e pesos dos atributos



Fonte: Elaborado pelo autor.

A avaliação da correlação entre os atributos tem objetivo de definir um ponto de partida para definição dos pesos, serão avaliados quais atributos possuem maior correlação entre si, nesses casos, serão definidos pesos menores.

A definição da quantidade de grupos é composta por três partes, são elas: análise descritiva dos atributos, aplicação de método estatístico de definição da quantidade ideal de grupos e definição hipóteses sobre os diferentes padrões de mobilidade com base na revisão. A análise descritiva dos atributos tem como objetivo identificar as características de heterogeneidade de cada um dos atributos, e, assim, permitir a definição de quantos grupos diferentes estes podem ser classificados.

A segunda etapa objetiva acrescentar ao processo decisório da quantidade de grupos uma ferramenta estatística (*Elbow method*), que indica a partir de que quantidade de grupos a variação dentro de cada grupo é estabilizada, portanto, identificando a quantidade de grupos que representam uma classificação natural do conjunto de dados analisado. Por fim, a última etapa, permite agregar ao processo decisório a quantidade de padrões esperados a partir dos resultados que foram encontrados na literatura.

Uma vez definida, o ponto de partida para esses pesos e a quantidade de grupos, serão identificados os padrões através do *k-means*. A etapa seguinte é de análise descritiva cujo propósito é avaliar se os pesos utilizados permitem que os padrões identificados sejam caracterizados e interpretados a partir de uma distribuição de frequência de cada atributo.

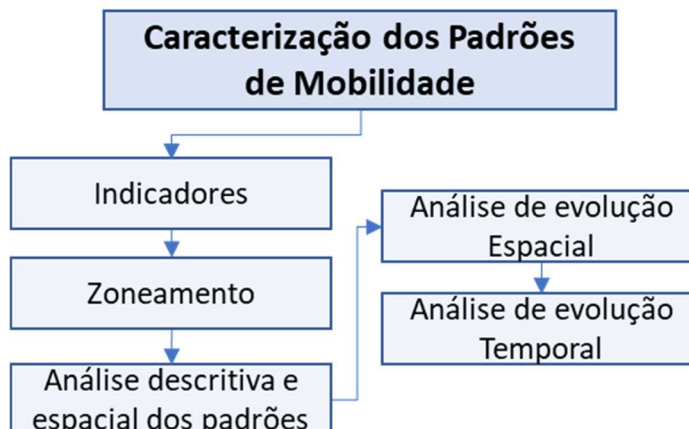
Quando necessário, serão realizados reajustes nos pesos, aumentando ou diminuindo, de forma arbitrária, mas procurando obter um resultado em que as distribuições de frequência dos atributos, para cada padrão, permitam de forma visual a sua caracterização, uma vez que os pesos foram ajustados serão novamente realizadas as etapas de identificação dos padrões e de análise descritiva. Essas três últimas etapas podem se repetir até que os resultados obtidos atendam as expectativas descritas.

5.4 Caracterização dos padrões de mobilidade

A etapa de caracterização é dividida em 6 partes: proposição de indicadores, definição do zoneamento, análise descritiva e espacial dos padrões, interpretação dos padrões, análises espaciais e, por fim, análises temporais. A sequência desta etapa pode ser visualizada na Figura 18. Na etapa de proposição de indicadores, serão relacionados os indicadores que serão utilizados, em cada uma das análises, a serem realizadas e quais os problemas de mobilidade que serão investigados por essa análise.

As etapas de análise descritiva dos padrões e distribuição espacial dos padrões darão subsídios para a etapa de interpretação dos padrões. As etapas de análise de evolução espacial e de evolução temporal serão responsáveis por avaliar como os padrões de mobilidade se alteraram entre os anos de 2014 e 2018. Todas as etapas de análises também serão responsáveis por investigar a relação dos padrões de mobilidade com problemas de mobilidade tipificados por Garcia *et al.* (2018).

Figura 18 – Metodologia de Caracterização dos Padrões de Mobilidade



Fonte: Elaborado pelo autor.

5.4.1 Indicadores

A finalidade dos indicadores, os quais serão propostos, é facilitar a compreensão dos diferentes padrões de mobilidade identificados e como esses padrões se relacionam com os problemas de mobilidade. O conjunto de indicadores (Tabela 5) e a ordem em que são propostos busca criar uma linha de raciocínio que facilite uma melhor compreensão dos resultados a serem realizados.

Inicialmente, são propostos indicadores que serão responsáveis pelas análises descritivas e espaciais dos padrões a fim de verificar as hipóteses distribuição espacial dos padrões e sua relação com o problema de distribuição desigual da mobilidade. Em seguida, são propostos indicadores que permitem avaliar a evolução desses padrões no tempo e no período de 2014 a 2018, que serão responsáveis por avaliar o agravamento dos problemas de distribuição desigual e insustentável da mobilidade. Estes últimos indicadores permitirão também avaliar como a queda na demanda está impactando cada padrão de mobilidade no espaço.

Tabela 5 – Lista de indicadores

Indicador	Análise	Tipo de problema	Hipóteses
Frequência média diária por zona			
Número de dias típicos de uso por zona	Análise descritiva e de distribuição espacial dos padrões	Distribuição desigual da mobilidade	(2)
Proporção dos padrões por zona			
Viagens produzidas em um dia / Domicílios de baixa) por zona			
Total de usuários por padrão	Análise de evolução temporal	Distribuição insustentável da mobilidade	
Diferença do indicador viagens produzidas por domicílios de baixa renda por zona			
Diferença de viagens por domicílios por viagens por domicílios 2014 por zona	Análise de evolução espacial	Distribuição desigual e insustentável da mobilidade	(3)
Diferença de viagens produzidas entre os anos por zona por zona			
Diferença de usuários entre os anos por zona			

Fonte: Elaborado pelo autor.

Os dados que permitiram a classificação de domicílios por renda serão buscados nos dados do Censo 2010 e serão classificados a partir do método aplicado por Sousa (2019) e Pinto (2020), em que foi definido o grupo de baixa renda para Fortaleza como sendo aquele composto pelos 60% domicílios de menor renda domiciliar. Os dados de domicílios de baixa renda utilizados por Sousa (2019) foram para o ano de 2015. Portanto, foi necessário estimar a quantidade de domicílios de baixa renda para os anos de 2014 e 2018, para isso foi utilizado o fator de crescimento calculado por Pinto (2020), que é de 1,02 para cada ano. Para calcular os indicadores, serão necessários somente os dados de domicílios do grupo de baixa renda.

Os indicadores que serão utilizados nas análises descritivas e espaciais dos padrões serão calculados com base na amostra de dados georreferenciadas listados na seção 5.2.3. Porém, os indicadores utilizados nas análises de evolução temporal e espacial levarão em consideração todos os usuários identificados na base de *Smart Card*, assim como as viagens do tipo inteira.

Fazendo-se necessário a realização de alguns ajustes para a utilização desses dados. Primeiramente, os usuários não encontrados nas bases de cadastros não permitem que sejam localizados os seus endereços de residência, impossibilitando a estimativa da origem das viagens através do método utilizado para os usuários cadastrados. Como solução, esses usuários serão

distribuídos nas zonas adotadas a partir da premissa de que eles possuem uma distribuição espacial e por padrão igual ao dos usuários cadastrados, utilizando a Equação 5 para calcular essas proporções.

$$PROP_{P,Z} = \frac{USUÁRIOS_CAD_{P,Z}}{USUÁRIOS_CAD_Z} \quad 5$$

As viagens do tipo “inteira” não permitem serem rastreadas, logo, não é possível identificar os padrões para os usuários que realizam essas viagens, portanto para distribuir essas viagens é necessário utilizar a mesma premissa dos usuários não cadastrados, além de considerar que os usuários que geram essas viagens possuem uma frequência média de utilização iguais aos dos usuários que tiveram os seus padrões identificados. Considerando-se essas duas premissas, é possível calcular quantos usuários foram responsáveis por gerar as viagens inteiras identificadas e então distribuí-los.

Os indicadores frequência média diária por zona e número de dias típicos de uso por zona (Equações 6 e 7) tem o propósito de analisar a distribuição espacial desses aspectos da mobilidade para cada padrão no zoneamento proposto. O indicador de frequência média busca identificar as regiões que produzem mais viagens por usuário em um dia útil, representando a característica de intensidade da mobilidade. O indicador de número de dias busca identificar as regiões que tem maior regularidade durante o ano para cada padrão. O indicador de proporção de usuários por padrão (Equação 8) busca identificar quais regiões cada padrão está mais presente.

O indicador de viagens por domicílios (Equação 9) é uma adaptação do indicador proposto por Henrique (2004) que tem calcula uma taxa de mobilidade para cada zona, possibilitando a identificação de regiões com maiores e menores valores dessa taxa. Neste trabalho este indicador será calculado de modo geral e para cada padrão. Esses indicadores têm como objetivo principal identificar problemas de distribuição desigual da mobilidade. Além dessas hipóteses, este trabalho também busca confirmar os resultados obtidos por Henrique (2004) para o indicador de taxa de mobilidade. Ou seja, que zonas periféricas, por estarem em zonas apresentam menores taxas de mobilidade. Henrique (2004) sugere que um dos motivos dessa baixa taxa de mobilidade nas regiões periféricas são os baixos níveis de acessibilidade.

$$I_FREQMED_{P,Z} = \frac{\sum_P A_FREQMED_{P,Z}}{USUÁRIOS_Z} \quad 6$$

$$I_NDIAS_{P,Z} = \frac{\sum_P A_NDIAS_{P,Z}}{USUÁRIOS_Z} \quad 7$$

$$I_PROP_{P,Z} = \frac{USUÁRIOS_{P,Z}}{USUÁRIOS_Z} \quad 8$$

$$I_VIAGENS_PDBR_{P,Z} = \frac{\sum_P VIAGENS_{P,Z} \times 100}{DOMICÍLIOS_BR_Z} \quad 9$$

Os indicadores diferenças de usuários, diferença de viagens e diferenças de viagens por domicílios (Equações 10 a 13) se propõem avaliar a evolução de cada padrão entre os anos de 2014 e 2018. Esses indicadores têm como objetivo principal identificar problemas de distribuição desigual e insustentável da mobilidade.

$$I_DIF_USUÁRIOS_P = USUÁRIOS_{2018,P} - USUÁRIOS_{2014,P} \quad 10$$

$$I_DIF_USUÁRIOS_{Z,P} = USUÁRIOS_{2018,Z,P} - USUÁRIOS_{2014,Z,P} \quad 11$$

$$I_DIF_VIAGENS_{Z,P} = (VIAGENS_{2018,Z,P} - VIAGENS_{2014,Z,P}) \quad 12$$

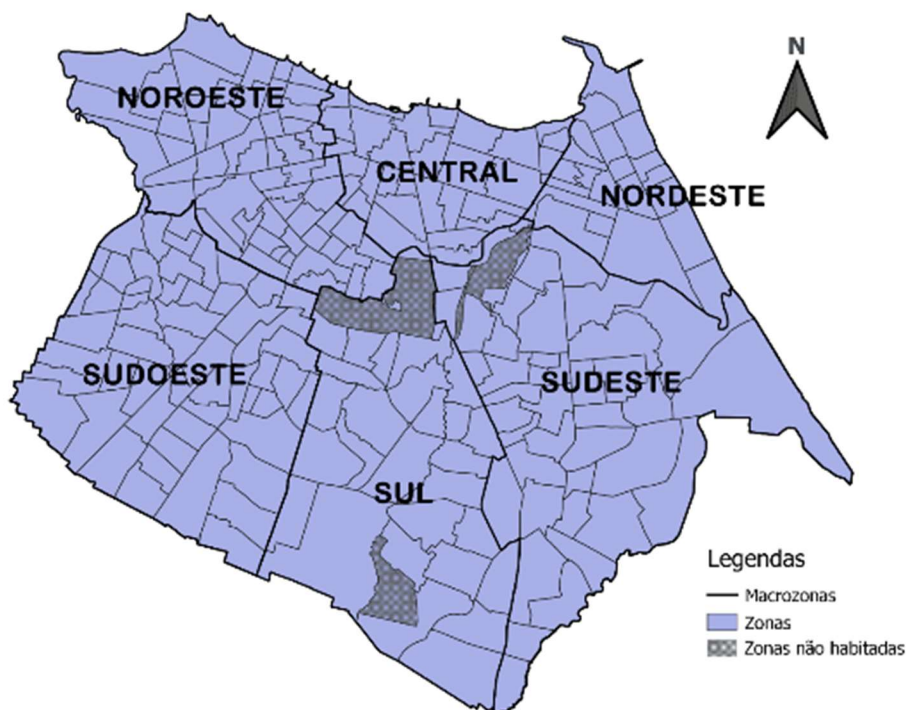
$$I_DIF_VIAGENS_PDBR_{Z,P} = (I_VIAGENS_PDBR_{2018,P,Z} - I_VIAGENS_PDBR_{2014,P,Z}) \quad 13$$

5.4.2 Zonamento

Uma vez que, os problemas de mobilidade se distribuem no espaço, faz-se necessária a utilização de um zoneamento no processo de caracterização. Para realizar as análises espaciais dos indicadores, adotou-se o zoneamento proposto por Lima (2017), que definiu as zonas de

Fortaleza a partir da agregação de setores censitários com características demográficas e socioeconômicas similares para a população usuária do transporte público (Figura 19).

Figura 19 – Proposta de Zoneamento para Fortaleza



Fonte: Adaptado de Lima (2017).

Esse zoneamento permite relacionar os padrões de mobilidade às características dos usuários em cada zona. Lima (2017) propõe uma divisão de Fortaleza em 241 zonas respeitando também as divisões por bairros, em que cinco foram descartadas por se tratar de áreas de parques e/ou corpos de água, além da zona na qual se encontra o Aeroporto Internacional Pinto Martins. Lima (2017) também divide Fortaleza em seis macrozonas (Central, Noroeste, Nordeste, Sudeste, Sudoeste).

Essas macrozonas são propostas através de uma consideração dos principais eixos viários e de aspectos sociais e econômicos dos bairros. A Região Central é composta por um grupo de zonas limitadas pelo primeiro anel viário da cidade, e compreendem o centro histórico, econômico, financeiro e administrativo da capital. As demais regiões, fazendo um contraponto a região central, serão chamadas de regiões periféricas. Sendo essa característica de periferia das zonas mais evidente à medida que estas se distanciam da região central.

Para o indicador de viagens por domicílios de baixa renda foram selecionadas e removidas das análises as zonas com menor quantidade de domicílios (23 zonas). Somados os domicílios de baixa renda dessas zonas representam 1% dos domicílios de baixa renda de Fortaleza. Essa decisão se deu devido a pouca quantidade de residências de baixa renda, o que poderia enviesar os resultados da análise e como os domicílios de baixa renda utilizados por Sousa (2019) são valores extrapolados e distribuídos para um novo zoneamento, esses valores podem não ser precisos para essas zonas específicas.

5.4.3 Análise descritiva e espacial dos padrões

Nesta etapa será analisada as seguintes hipóteses: 1) Diferentes padrões de mobilidade são esperados para os usuários de transporte público por ônibus em Fortaleza; 2) Os padrões de mobilidade que apresentam uso dependente e mais e regular são proporcionalmente mais presentes em zonas mais periféricas e usuários pertencentes a padrões de uso não dependente e menos regular residem em zonas mais centrais. Esse fenômeno mostra o quão dependente do transporte coletivo podem ser os usuários de baixa e média renda que vivem em zonas periféricas.

Inicialmente, as análises dos indicadores (Equação 6 a 9) se propõem a investigar como a mobilidade dos padrões é distribuída no espaço através da utilização dos atributos responsáveis por identificar os padrões de mobilidade como indicadores de mobilidade. Para isto serão analisadas as distribuições espaciais, o *Lisa Cluster Map* e um *Boxplot* desses indicadores para cada padrão de modo a avaliar a dependência espacial dos indicadores e para identificar a formação de *clusters*. Em seguida será analisado como as zonas são compostas por cada um dos padrões identificados através de um indicador de proporção. Espera-se com estas análises representar as relações entre os padrões de mobilidade com os problemas de mobilidade do tipo distribuição desigual. As análises propostas foram realizadas nos softwares R Studio, QGIS e GEODA.

De modo a identificar se algum grupo de usuários específicos de algum tipo de cartão deixaram de utilizar o sistema, verificou-se como a composição dos grupos por tipo de cartão (Vale Transporte, Estudante e Gratuidade) se altera ao longo dos anos analisados (2014-2018), o propósito é identificar qual tipo de cartão mais afetado pela queda na demanda.

5.4.4 Análise da evolução temporal

Nesta etapa, busca-se identificar se algum grupo de usuários específicos de um dos tipos de cartão deixou de utilizar o sistema, verificou-se como a composição dos grupos por tipo de cartão (Vale Transporte, Estudante e Gratuidade) se altera ao longo dos anos analisados (2014-2018), o propósito é identificar qual tipo de cartão mais afetado pela queda na demanda.

Para analisar a dinâmica dos padrões ao longo do tempo, conforme proposto por Briand *et al.* (2017), foi gerado um diagrama de Sankey com a diferença de usuários (Equação 11) entre os padrões para os anos de 2014 e 2018. Essa análise avalia a dinâmica dos usuários e padrões ao longo do tempo, podendo-se identificar quais usuários entraram e deixaram o sistema, bem como se estes alteraram o seu padrão de mobilidade entre os anos. Espera-se com esses indicadores, avaliar quais os padrões de mobilidade são mais afetados pelo problema de queda na demanda para os anos de 2014 e 2018.

5.4.5 Análise da evolução espacial

Nesta etapa será avaliada a distribuição espacial de três indicadores: diferença de usuários entre os anos por zona (Equação 12), diferença de viagens produzidas entre os anos por zona (Equação 13) e diferença do indicador de viagens por domicílios de baixa renda entre os anos por zona (Equação 14). Espera-se com estes indicadores avaliar quais os padrões e quais zonas foram mais afetadas pela queda na demanda entre o período analisado, investigando assim as hipóteses: 4) Os padrões de mobilidade de uso dependente e mais regular são mais impactados pela queda na demanda, implicando o agravamento dos problemas de distribuição desigual e insustentável da mobilidade ao longo dos anos; 5) Padrões de mobilidade menos regulares são menos impactados pela queda na demanda, implicando também o agravamento dos problemas de distribuição desigual e insustentável da mobilidade ao longo dos anos. Nessa etapa também serão analisadas as distribuições espaciais, o Lisa Cluster Map e um Boxplot dos indicadores para cada padrão a fim de investigar a dependência espacial e a formação de clusters espaciais desses indicadores.

6 ANÁLISES E DISCUSSÕES DOS RESULTADOS

Neste capítulo, serão apresentados os resultados do método de identificação e caracterização dos padrões de mobilidade para os usuários de transporte público por ônibus de Fortaleza utilizando os dados de *Smart Card*. Assim como o método os resultados serão divididos em duas macro-etapas: Identificação dos padrões e Caracterização dos padrões.

6.1 Identificação dos padrões de mobilidade

A literatura permitiu identificar alguns tipos de grupos com relação ao padrão temporal: i) usuários pendulares, que usam o sistema com propósito de acessar atividades de trabalho ou educação em horários de pico e usuários pendulares tem um uso constante do sistema; ii) usuários que apresentam um uso temporal irregular (não usam nos mesmos horários ao longo da semana mas possivelmente acessam as mesmas atividades), mas com um uso frequente do sistema; e iii) usuários com uso temporal e frequência de utilização irregular (usuários esporádicos). Percebe-se, que esses grupos são formados por uma combinação entre dois fatores principais: Regularidade de uso e Intensidade de uso. A intensidade de uso se refere ao número viagens que um usuário realiza no sistema em um dado dia útil. A regularidade de uso do sistema se refere a quantidade de vezes que o padrão é repetido ao longo do ano. Porém, nesta pesquisa, também será considerado a duração da atividade através do atributo de distância temporal médio e do desvio padrão dessa distância temporal, que, além de remeter ao tipo de atividade realizada, por vezes, também pode ser um indicativo de regularidade.

6.1.1 Análise descritiva dos atributos

Para identificar os padrões de mobilidade através do *k-means* foi necessário definir a quantidade de grupos esperados e os pesos para cada atributo. A primeira análise realizada que irá guiar na definição desses parâmetros é uma análise descritiva dos atributos para os anos de 2014 e 2018. No processo de identificação dos padrões, os dados de 2014 e 2018 serão unificados. Assim, é realizado uma análise descritiva da distribuição da frequência relativa de cada atributo para os dois anos considerando que os dois anos formam um único conjunto de dados.

O resultado (Figura 20 e Tabela 6) dessas análises permitem que alguns aspectos importantes sejam destacados: 1) As principais diferenças entre os anos de 2018 e 2014 são de usuários que possuem valores mais elevados de cada atributo; 2) Os atributos de desvio da frequência média, distância temporal médio e número de dias do ano possuem um maior coeficiente de variação; 3) O atributo de distância temporal média tem uma característica de distribuição achatada, sendo os seu pico muito próximo das classes mais próximas; 4) O atributo número de dias do ano possui um pico elevado a esquerda, mostrando que há uma grande quantidade de usuários que utilizam o sistema com pouquíssima frequência ao longo do ano, além disso, o resto da distribuição apresenta um achatamento para direita ao longo das classes.

Três dos atributos apresentaram uma maior variação: Desvio da frequência média, Distância temporal média e Número de dia de uso durante o ano. Essa variação mais elevada significa que implica em uma maior heterogeneidade dos usuários para esses atributos em comparação ao atributo de frequência média e de desvio da distância temporal. A baixa variação da frequência média é esperada, uma vez que a heterogeneidade e os valores para esse evento são mais previsíveis. A variação para o atributo de desvio da frequência média indica que, apesar do evento de realizar viagens (frequência média) seja previsível, a quantidade das viagens em um dia útil pode variar dependendo do dia da semana, ou do mês. Essa variação pode ocorrer pela escolha de utilização de outros modos, ou pela realização de uma atividade que não possua um padrão regular.

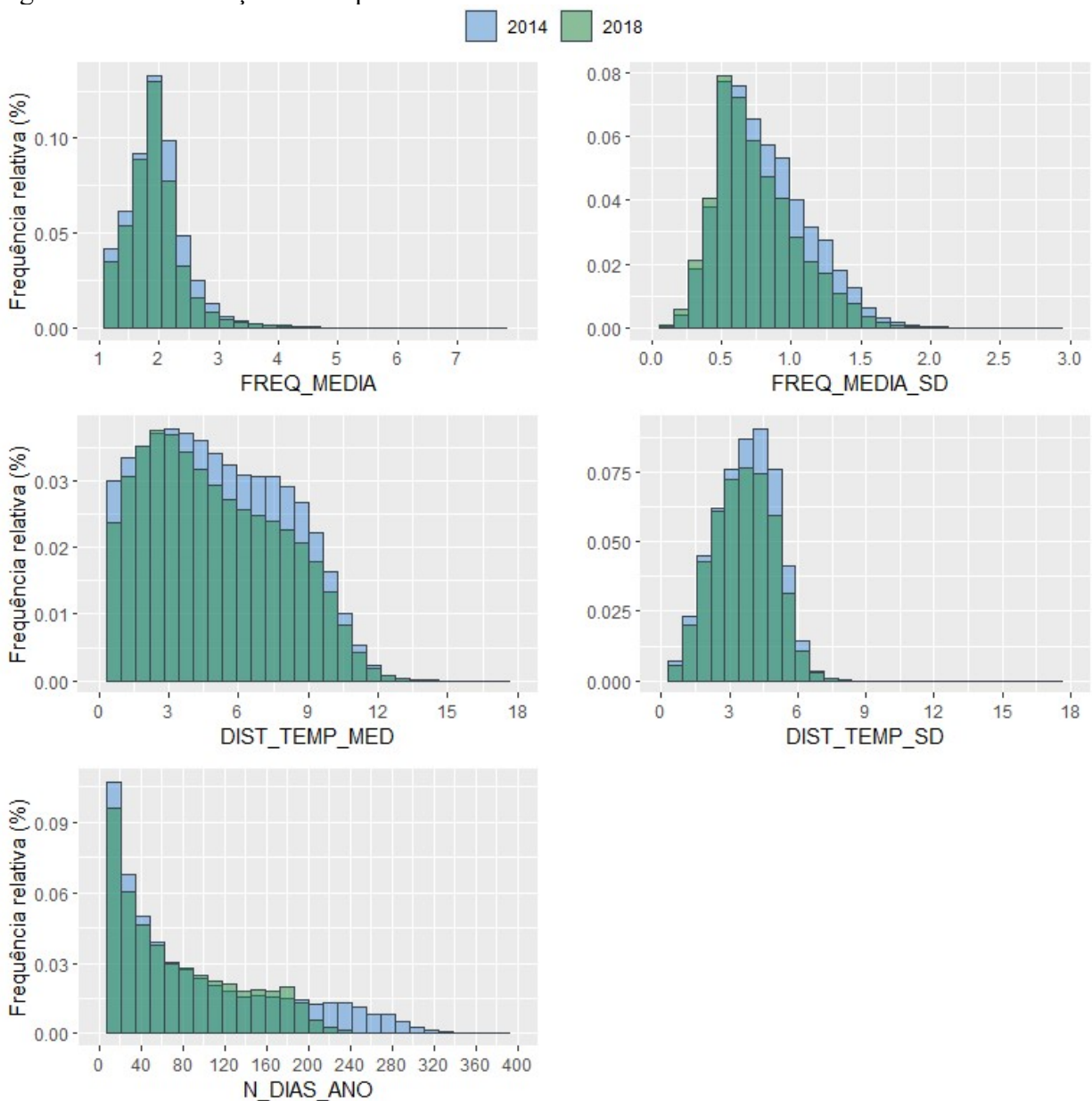
Uma vez que se percebe uma variação significativa dentro de cada atributo, essas distribuições, medidas de tendência e de variação corroboram com hipótese de que diferentes padrões de mobilidade são esperados para os usuários de transporte público por ônibus de Fortaleza e justificam a necessidade de categorizar esses usuários em a partir do seu padrão.

Tabela 6 – Medidas de tendência central e variação

Atributo	2014				2018			
	Média	Mediana	Desvio Padrão	Coefficiente de variação	Média	Mediana	Desvio Padrão	Coefficiente de variação
FREQ_MEDIA	1.95	1.94	0.48	0.25	1.92	1.91	0.48	0.25
FREQ_MEDIA_SD	0.80	0.75	0.32	0.40	0.74	0.69	0.31	0.42
DIST_TEMP_MED	4.96	4.73	2.97	0.60	4.83	4.45	2.91	0.60
DIST_TEMP_SD	3.64	3.75	1.39	0.38	3.57	3.62	1.33	0.37
NDIAS	93.10	64.00	80.73	0.87	75.95	59.00	58.85	0.77

Fonte: Elaborado pelo autor.

Figura 20 – Distribuição de frequência relativa dos atributos



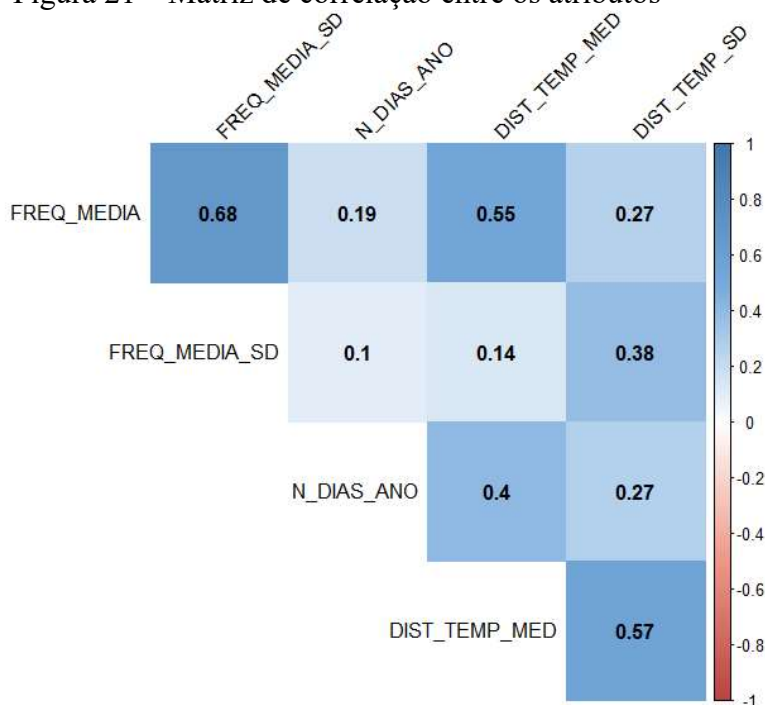
Fonte: Elaborado pelo autor.

6.1.2 Análise de correlação dos atributos

Em seguida é calculado a correlação entre cada atributo (Figura 21), com objetivo de identificar se os atributos calculados possuem algum grau de relação entre si. Esta análise permitirá justificar algumas escolhas para a determinação dos pesos dos atributos. Os valores mais elevados de correlação já eram esperados, uma vez que a frequência média e a e o desvio da

frequência média, são calculadas a partir dos valores destes indicadores para cada dia útil que o usuário realizou viagens. O mesmo acontece para a distância temporal média e o seu desvio padrão. Também era esperado que a correlação entre a frequência média e a distância temporal média, uma vez que usuários que realizam mais viagens ao longo dia tendem a possuir atividades mais longas. Embora exista correlação entre os atributos, considera-se que ainda é possível combiná-los para identificar padrões.

Figura 21 – Matriz de correlação entre os atributos



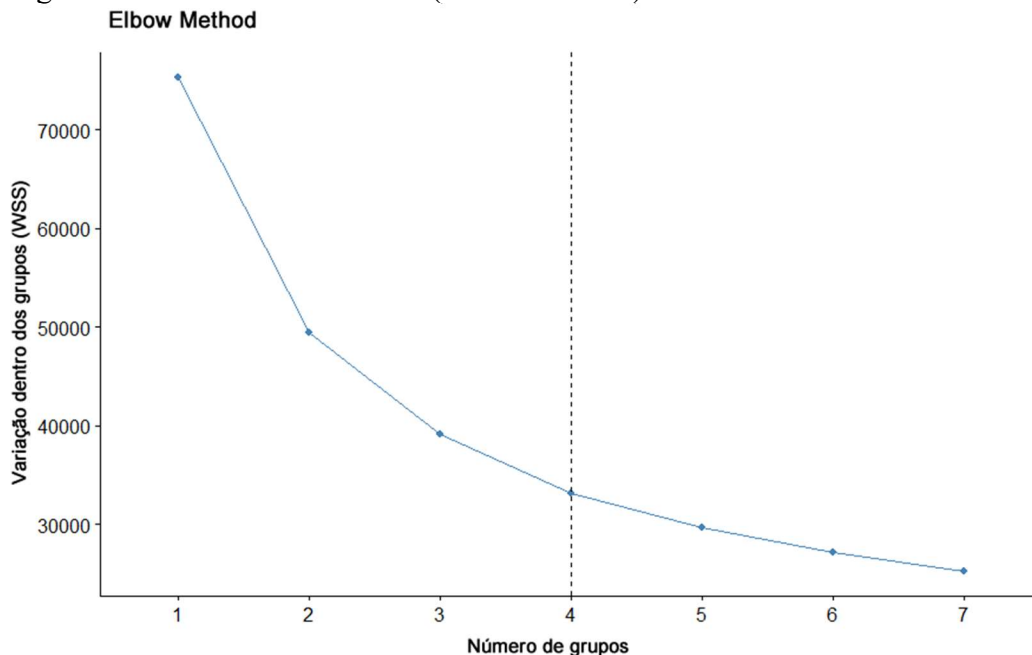
Fonte: Elaborado pelo autor.

6.1.2 Aplicação de método estatístico para definição de grupos

A etapa seguinte é aplicação de um método estatístico para definição a quantidade ideal de grupos. O método do cotovelo (*Elbow Method*) mede a variação dentro de cada *cluster* para diferentes quantidades de agrupamentos e utiliza essa medida como indicador para sugerir a quantidade ideal de grupos. É executado o *k-means* para N (quantidade de grupos) de 1 a 7, considerando o mesmo peso para todos os atributos, e então é calculado a variação dentro dos grupos. A definição da quantidade ideal de grupos é realizada de forma visual, identificando em que ponto do gráfico a queda da variação começa a se estabilizar. O resultado (Figura 22) deste

método mostrou que a partir de N igual a 4 (quatro grupos), a queda da variação começa a diminuir consideravelmente.

Figura 22 – Método do cotovelo (Elbow Method)



Fonte: Elaborado pelo autor.

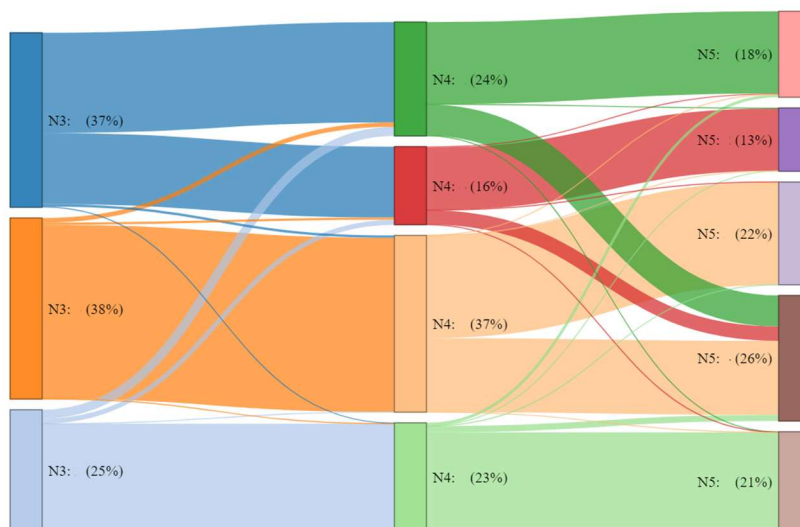
Para avaliar a distribuição dos usuários entre os cluster assim como o formato desses clusters para diferentes números de grupos, foi gerado um Diagrama de Sankey. O diagrama de Sankey é uma representação visual de um fluxo. O fluxo envolve uma transferência de algum objeto em análise de uma etapa para outra e a utilização desse diagrama tem objetivo de facilitar o senso de interpretação desse tipo de informação. Essas transferências são representadas por linhas cujas espessuras são proporcionais ao volume do objeto que está sendo transferido.

O resultado dessa análise (Figura 23), permitiu visualizar como os usuários se distribuem se a quantidade de grupos fosse aumentada. Foi selecionado o intervalo de N igual 3 até N igual a 5, que é o intervalo que o *Elbow Method* tem a sua estabilização percebida. De N igual 3 para N igual a 4, dois grupos permanecem praticamente os mesmos, enquanto um grupo é dividido, formando dois novos grupos. Essa dinâmica sugere que existe 4 *clusters* bem formados.

De N igual 4 para N igual a 5, um novo grupo é formado, esse novo grupo recebe usuários com mais intensidade de três grupos. Essa divisão sempre existirá, e sempre poderá ser

observada, uma vez que o *k-means* requer como entrada o número de grupo e sempre irá alocar as observações de acordo com o número de grupos definidos. Porém, as dinâmicas observadas de **N** igual **3** para **N** igual a **4** são mais naturais, enquanto de **N** igual **4** para **N** igual a **5**, parecem ser mais um rearranjo dos usuários.

Figura 23 – Distribuição dos usuários por padrão para N=3, 4 e 5



Fonte: Elaborado pelo autor.

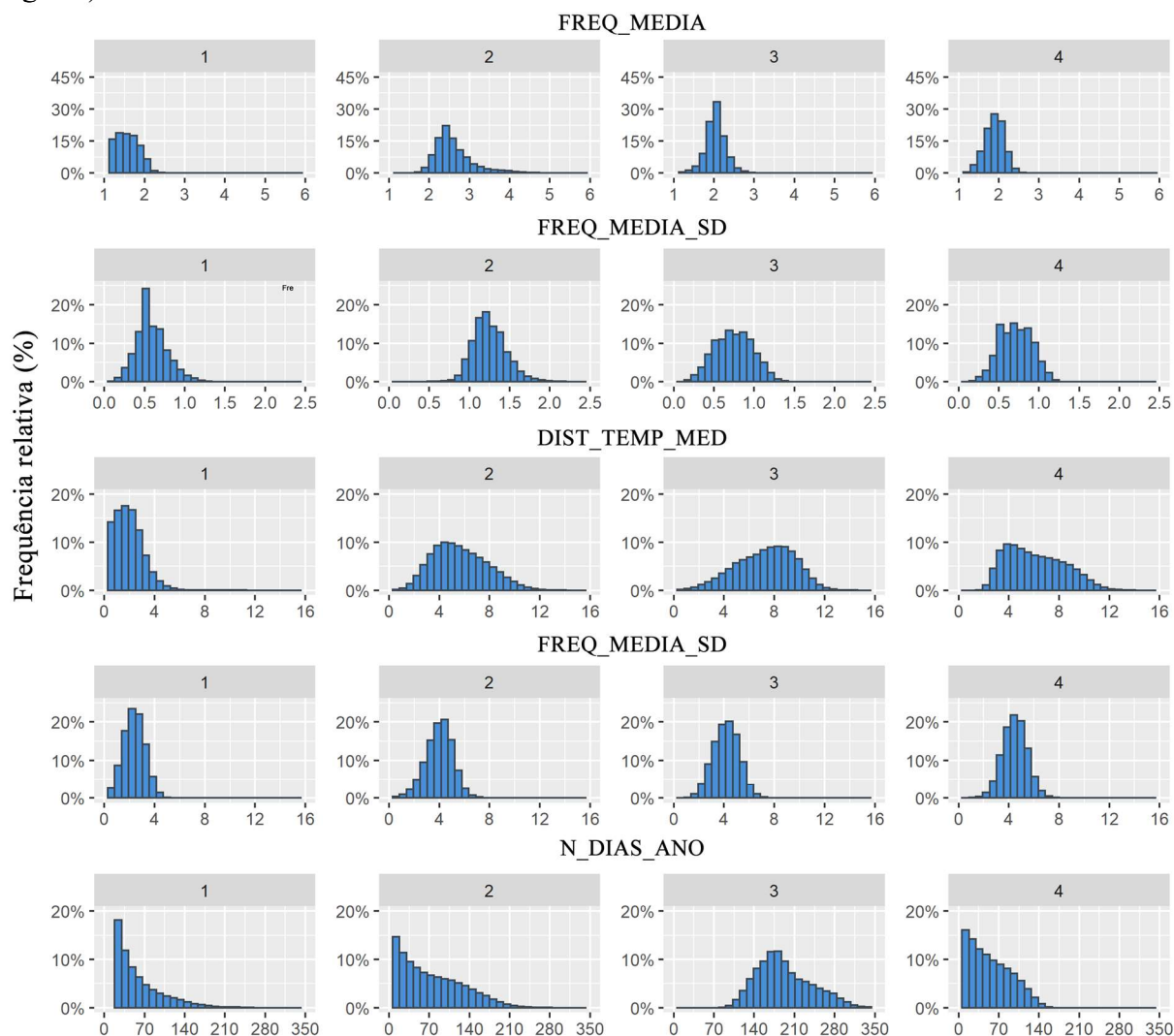
A literatura demonstrou que três grupos majoritários podem ser observados para esse fenômeno, porém, indo além dos grupos sugeridos pela literatura e considerando os usuários do transporte público de Fortaleza e os atributos definidos para representar a mobilidade destes usuários, o método estatístico e a análise de composição e fluxo dos grupos foi definido que a quantidade de grupos esperado neste trabalho será 4.

6.1.4 Definição dos pesos

Inicialmente foi executado o *k-means* sem definir pesos para nenhum dos atributos e então avaliado a distribuição de frequência relativa de cada atributo para cada um dos padrões, ou seja, o peso de cada atributo inicialmente é 1. A partir dessas distribuições (Figura 24), é possível perceber que os atributos de distância temporal média e número de dias do ano possuem

distribuições de frequência mais dispersas, dificultando, de certa forma, a interpretação dos padrões.

Figura 24 – Distribuições das frequências relativas dos atributos para cada padrão (pesos iguais)



Fonte: Elaborado pelo autor.

Levando em consideração o resultado destas distribuições, e o resultado da matriz de correlação dos atributos, que mostraram que os atributos de frequência média, desvio padrão da frequência média e desvio padrão distância temporal possuem maior níveis de correlação, optou-se por aumentar de 1 para 2 os pesos dos atributos de distância temporal média e número de dias. Na figura 24, que se encontra na próxima seção, percebeu-se uma leve mudança no formato das

distribuições dos atributos distância temporal média e número de dias do ano, mas que permitiram representar melhor esses padrões visualmente. O ajuste de pesos não parou nesta etapa, foram realizados mais alguns testes, onde foi aumentado ainda mais os pesos dos atributos de distância temporal média e número de dias, mas os resultados não apresentaram melhoras relevantes na visualização das distribuições. Por tanto, optou-se por definir o número definitivo dos pesos em 1 para frequência média, 1 para desvio da frequência média, 2 para distância temporal média, 1 para desvio da distância temporal média e 2 para número de dias.

6.2 Caracterização dos padrões de mobilidade

6.2.1 Análise descritiva dos padrões

Uma vez definido os atributos, a quantidade de grupos esperadas e os pesos de cada atributo foi possível executar o *k-means* para identificar os grupos. Cada grupo representará um padrão de mobilidade. Como análise inicial para esses padrões, foi avaliado a distribuição de frequência relativa para cada atributo (Figura 24) e uma análise descritiva por tipo de cartão para cada cluster (Tabela 7).

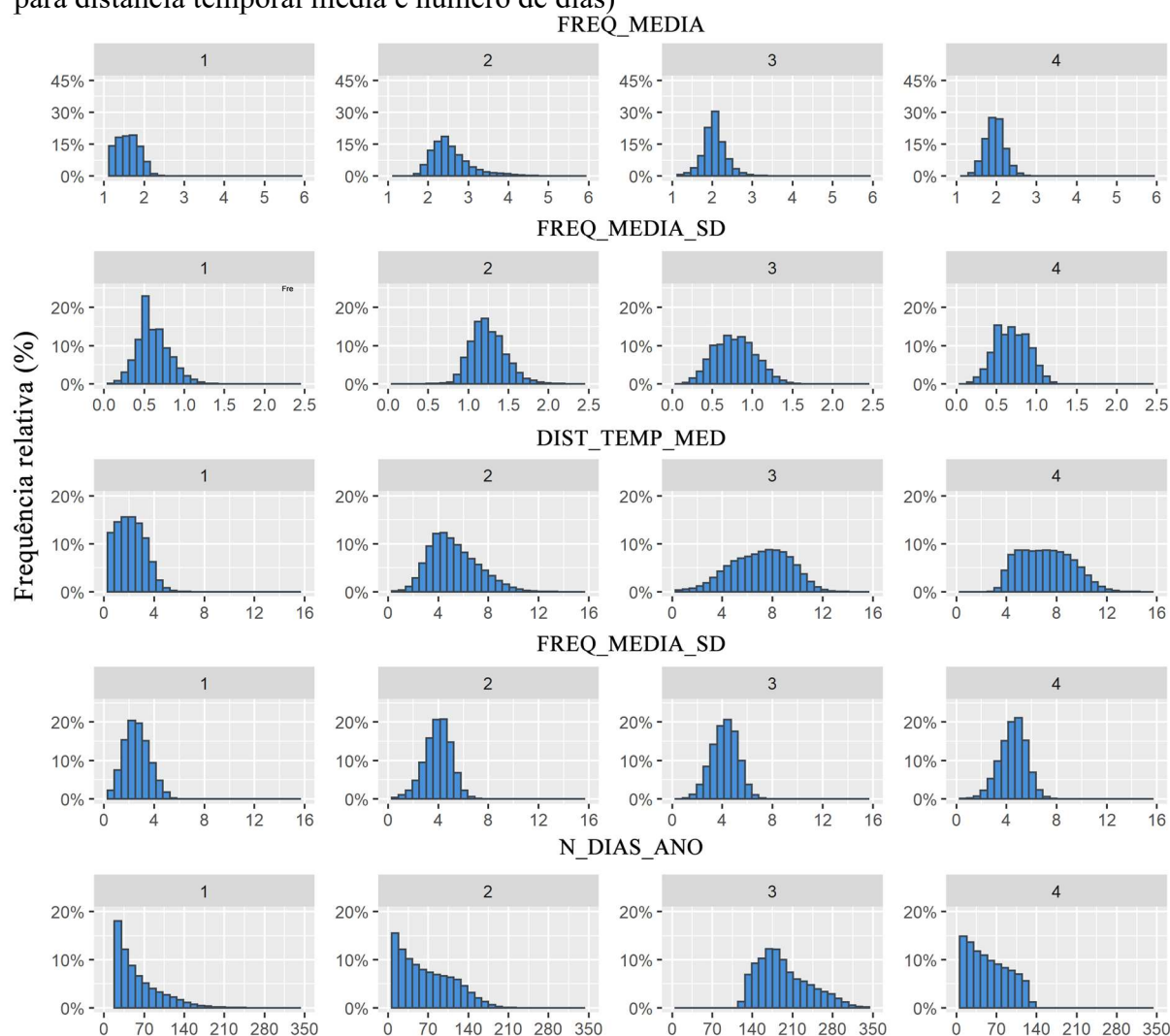
Tabela 7 – Análise descritiva dos clusters

Padrão	Tipo de Cartão	Quantidade de usuários (2014 +2018)	Proporção dentro do padrão	Média das frequências	Média das distâncias temporais entre validações	Média do número de dias
1	Estudante	357,666	51%	1.6	2.3	43
	Gratuidade	123,918	18%	1.5	1.3	40
	Vale transporte	217,081	31%	1.5	1.8	49
	Total	698,665		1.6	2.0	44
2	Estudante	49,879	16%	2.3	5.1	66
	Gratuidade	32,647	11%	3	4.3	76
	Vale transporte	226,089	73%	2.6	5.4	66
	Total	308,615		2.6	5.2	67
3	Estudante	106,515	25%	2.1	6.2	181
	Gratuidade	9,232	2%	2.2	5.3	196
	Vale transporte	312,027	73%	2	7.4	198
	Total	427,774		2.0	7.1	194
4	Estudante	128,294	28%	1.9	6.1	61
	Gratuidade	3,829	1%	1.9	5.7	64
	Vale transporte	318,208	71%	2	7.6	56
	Total	450,331		2.0	7.2	58

Fonte: Elaborado pelo autor.

A análise dos atributos para cada grupo (Figura 25 e Tabela 7) permitem identificar os três padrões sugeridos pela literatura (Usuários regulares e dependentes do sistema; Usuários dependentes do sistema, mas não regular; e usuários esporádicos). Porém, o que os resultados indicam, é que o padrão de usuários dependente, mas não regulares (Padrões 2 e 4), podem ser classificados em dois diferentes padrões (fenômeno também visualizado na Figura 23).

Figura 25 – Distribuições das frequências relativas dos atributos para cada padrão (pesos 2 para distância temporal média e número de dias)

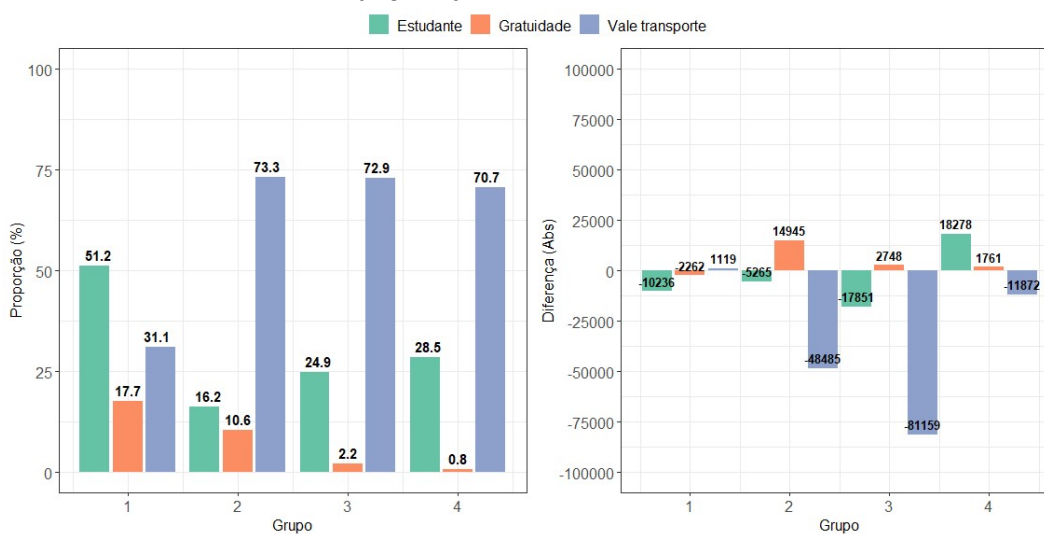


Fonte: Elaborado pelo autor.

Em seguida, é avaliado a distribuição dos tipos de cartão por padrões e aproveitou-se para apresentar a análise de evolução temporal por tipo de cartão entre 2014 e 2018 (Figura 26). Dessas análises, alguns aspectos merecem ser destacados: 1) Os padrões 2, 3 e 4 são formados

majoritariamente por usuários de cartão do tipo Vale transporte; 2) O padrão 1 é formado majoritariamente por estudantes; 3) Os usuários que utilizam cartão do tipo Gratuidade pertencem majoritariamente aos padrões 1 e 2; 4) Os usuários que mais deixaram de utilizar o sistema foram os usuários de Vale Transportes dos padrões 2 e 3; 5) Houve um aumento de gratuidade entre os anos de 2018 e 2014; 6) Houve uma queda na quantidade de estudantes para os padrões 1, 2 e 3, e um aumento de estudantes para o padrão 4, porém a diferença de todos os padrões resulta em um valor negativo de 15.074 estudantes.

Figura 26 – Proporções por tipo de cartão de cada grupo (2014 e 2018), e diferença de usuários entre os anos 2018 e 2014



Fonte: Elaborado pelo autor.

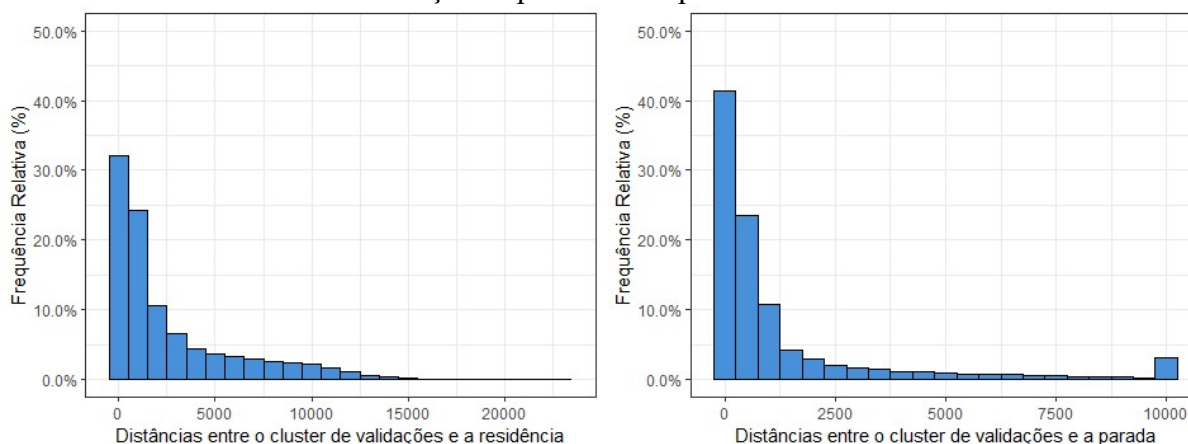
Esses resultados corroboram com a análise dos padrões identificados. E, conforme já esperado, sugere que os usuários que utilizam vale transporte tendem a ser mais regulares do sistema. Mostra também que existe uma parcela de estudantes que utiliza o sistema esporadicamente que provavelmente utiliza o sistema para realizar atividades outros e não do tipo estudo. Para a evolução dos tipos de cartão entre os anos 2014 e 2018, é possível notar que as maiores queda de usuários é para cartões do tipo vale transporte do padrão regular e dependente (Padrão 3), seguido do padrão regular e não dependentes (Padrão 2) que provavelmente realiza atividade irregular (trabalho informal ou atividade do tipo outros).

6.2.2 Análise descritiva dos padrões

A seguir, serão analisadas as distribuições espaciais dos indicadores propostos na metodologia para caracterizar os padrões identificados. As análises foram realizadas de forma separadas para os anos de 2014 e 2018, porém, como há indicadores e análises propostos especialmente para avaliar a evolução desses padrões e como o resultado dos padrões para os dois anos é muito parecido, inicialmente só serão apresentadas as análises para o ano de 2018 e a análises realizadas do ano de 2014 serão adicionadas ao apêndice.

Para analisar espacialmente esses dados, o local de origem das viagens geradas será o endereço de cadastrado presente na base de cadastro do bilhete único. Como forma de validar a utilização do endereço de cadastro como endereço de origem das viagens foram avaliadas as distâncias entre o *cluster* das primeiras validações dos usuários para os seus endereços de cadastro e as distâncias entre o *cluster* das primeiras validações e a parada mais próxima do endereço de cadastro (Figura 27).

Figura 27 – Distribuições das distâncias dos clusters de validações aos endereços de residência e distâncias do clusters de validações a parada mais próxima das residências



Fonte: Elaborado pelo autor.

Em média, 60% dos usuários possuem endereço de cadastro até 3 km do centroide do cluster de validações. Acontece que, em 2018, o sistema de validação dos ônibus em Fortaleza permitia que os usuários pudessem validar durante todo o percurso, por tanto, optou-se também, por identificar a parada mais próxima da residência que faça parte da linha frequente do usuário e avaliar a distribuição dessas distâncias. Para essa distribuição, em média 75% dos usuários

possuem endereço de cadastro com distância de até 1,5 km para uma parada de ônibus pertencente a sua linha frequente.

Apesar de existir alguns usuários com distâncias elevadas entre o cluster de validação e a parada mais próxima a residência para o endereço de residência, acredita-se que boa parte desses usuários utilizam algum outro modo para se deslocar até a parada. Por tanto, os resultados indicam que é possível a utilização do endereço de cadastro como origem da viagem.

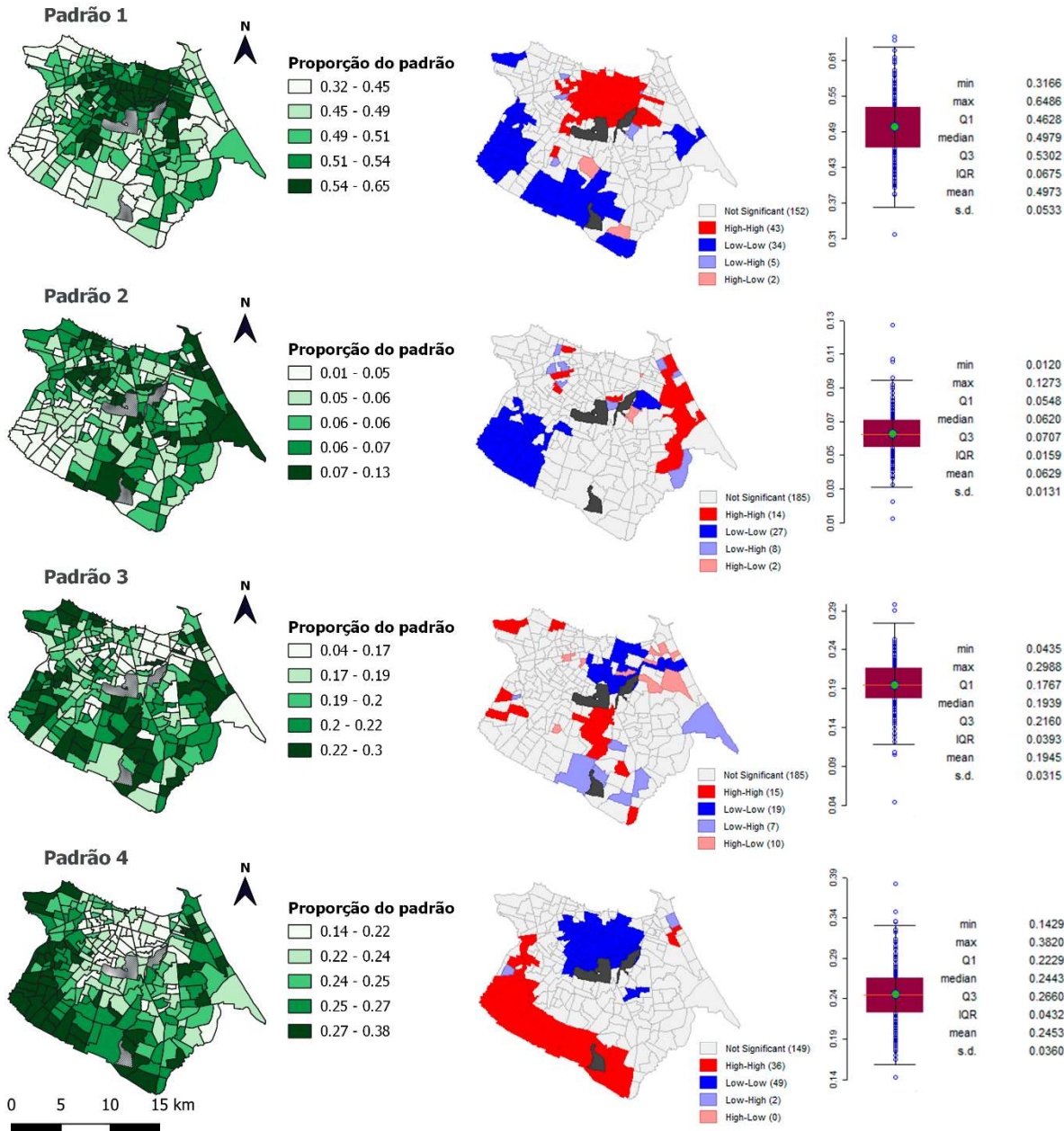
Uma vez definido que os endereços de cadastros serão a origem das viagens foram calculados os indicadores a partir da metodologia proposta. O primeiro indicador avaliado foi avaliado a proporções dos padrões por zona.

A distribuição espacial (Figura 28) mostra que o padrão de usuários esporádicos (padrão 1) está proporcionalmente mais presente nas zonas localizadas na região central, e essa proporção cai à medida que as se aproxima das zonas periféricas. Para os padrões de usuários dependentes (padrões 2, 3 e 4), essa distribuição ocorre de maneira inversa ao padrão 1, sendo proporcionalmente mais presentes nas zonas periféricas.

As análises de Lisa Map apresentam *clusters High-High* bem definidos nas regiões centrais para o padrão 1, e na região nordeste para o padrão 2 e nas regiões sul, sudeste e sudoeste para o padrão 4. É possível identificar também *clusters Low-Low* bem definidos nas regiões sudoeste para os padrões 1 e 2, na região central/nordeste para o padrão 3 e região central para o padrão 4.

Esses resultados corroboram com as hipóteses de que os padrões de mobilidade de uso dependente e mais regular são proporcionalmente mais presentes em zonas mais periféricas e menos presentes em zonas centrais. Os resultados permitem também verificar uma distribuição bem diferente para os padrões 2 e 4, que são de usuários dependentes, mas não regulares, e que provavelmente se diferenciam pela atividade realizada. O padrão 2 apresenta uma distribuição mais uniforme entre as zonas, porém com uma ausência bem concentrada na região sudoeste, diferentemente do padrão 4 que apresenta uma concentração elevada em toda a região sul.

Figura 28 – Mapa de quintis; Lisa Map e Boxplot para o indicador proporções dos padrões no ano de 2018



Fonte: Elaborado pelo autor.

Com os resultados obtidos até aqui, algumas interpretações sobre os padrões podem ser realizadas. O padrão de usuários mais esporádico (Padrão 1) está proporcionalmente mais presente nas zonas centrais de Fortaleza. A sua distribuição por tipo de cartão mostra que ele é composto majoritariamente (52%) por usuários de cartão do tipo Estudante. Esses estudantes possuem uma frequência média de 1,6 viagens por dia, uma distância temporal entre validações

média é de 2,3 horas e a média do número de dias é 43. Esses valores não se alteram muito para os outros tipos de cartão, apenas para o atributo de distância temporal onde há uma leve redução. Uma vez que as atividades do tipo estudo levam em torno de 4 a 6 horas para serem realizadas, a hipótese é que esse tipo de padrão, majoritariamente, realize viagens para realizar atividades do tipo outros. E como a média de número de dias de uso durante o ano é baixa, esses usuários não tendem a ser cativos do sistema de transporte público por ônibus. Essas características caracterizam este padrão como usuários esporádicos do sistema.

O padrão 2 tem pouca representatividade por zona, sendo mais presente em algumas zonas da região nordeste e menos presente nas zonas da região sudoeste. É formado em sua maioria por usuários que utilizam cartão do tipo Vale transporte (73%) que possuem uma frequência média considerada elevada (2.6 viagens por dia), uma distância temporal de média de 5,4 horas e número de dias de uso durante o ano com uma média de 66. Espera-se que uma atividade de trabalho formal possua pelo menos 6 horas de duração, por tanto, a hipótese é que os usuários deste padrão realizem atividades do tipo trabalho informal ou atividades do tipo outros. O número de dias de uso ao longo do ano mostra também que esses usuários não são completamente cativos do sistema, ou que a informalidade do tipo de atividade não requer uso todos os dias da semana.

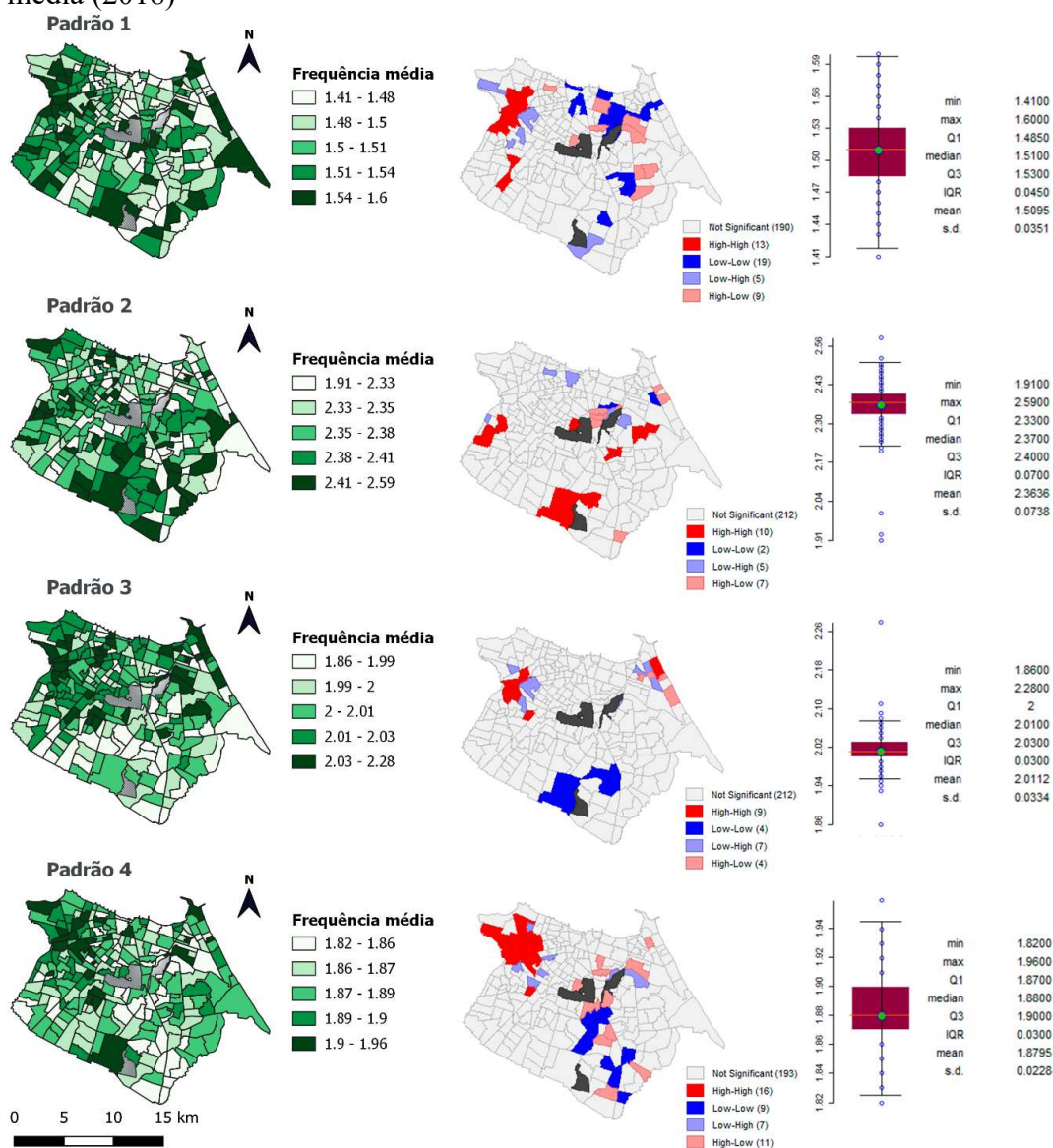
Já o padrão 3 é proporcionalmente mais presente nas regiões mais periféricas e menos presente nas zonas centrais. É formado majoritariamente também por usuários de cartão do tipo Vale transporte (73%), apresentando para estes usuários uma frequência média de 2 viagens por dia com uma distância temporal de 7,4 horas e uma média de número de dias de uso durante o ano de 198. A hipótese é que estes são os usuários cativos do sistema e que realizam atividades do tipo trabalho formal ou educação.

O padrão 4 apresentou uma presença elevada nas zonas periféricas, com uma concentração muito forte em toda a região sul e uma baixa concentração nas zonas centrais. Também é formado em sua maioria por usuários que utilizam cartão do tipo Vale transporte (71%). Os valores dos atributos foram muito parecidos com o do padrão 2, diferenciando-se somente na distância temporal média e na frequência média. Por tanto, a hipótese é de que esses usuários não são cativos do sistema e que realizam atividades do tipo trabalho formal ou educação.

Em seguida, foi avaliado a distribuição espacial do indicador de frequência média diária por padrão de mobilidade (Figura 29). Para todos os padrões percebe-se que o indicador

aumenta à medida que se aproxima das zonas mais periféricas. Também é possível perceber para todos os padrões que existe uma tendência de queda para o indicador à medida que as zonas se aproximam dos principais eixos viários e dos terminais de ônibus, que pode ser justificado pela menor necessidade de realização de integração por parte desses usuários. Esses resultados mostram que mesmo para usuários de um mesmo padrão, é possível identificar uma diferença na distribuição espacial da mobilidade, e que os usuários com maior intensidade de uso estão mais presentes em zonas periféricas.

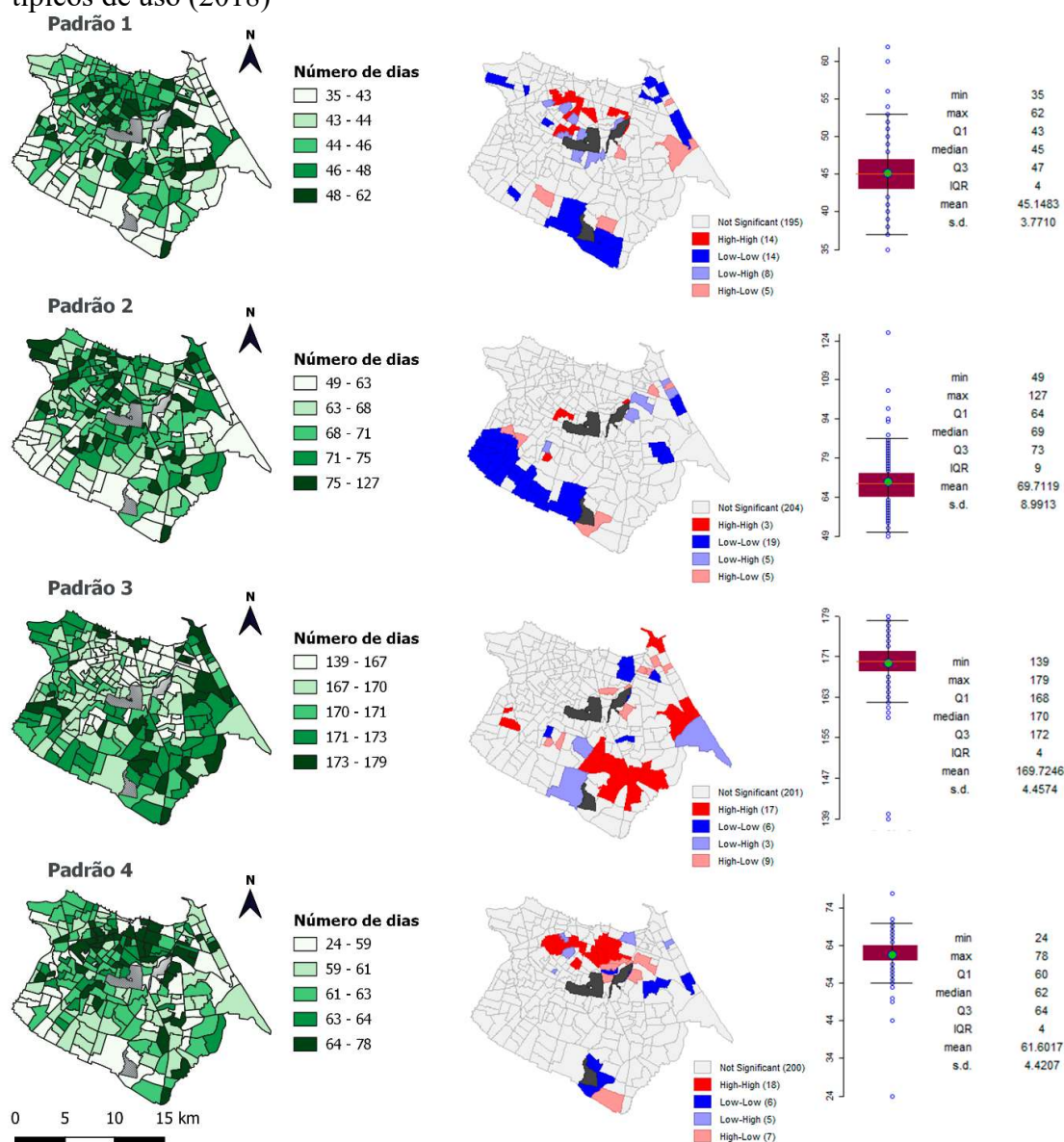
Figura 29 – Mapa de quintis; Lisa Map e Boxplot para o indicador de frequência média (2018)



Fonte: Elaborado pelo autor.

Analisando-se a distribuição espacial do indicador número de dias típicos de uso por zona (Figura 30), percebe-se também, que existe uma diferença na distribuição espacial deste indicador. Porém, neste indicador, para os padrões esporádicos e não dependentes do sistema (1, 2 e 4) os valores aumentam à medida que se aproxima das zonas mais centrais. E para o padrão de uso dependente e mais regular, o indicador aumenta à medida que se aproxima das zonas periféricas.

Figura 30 – Mapa de quintis; Lisa Map e Boxplot para o indicador de número de dias típicos de uso (2018)

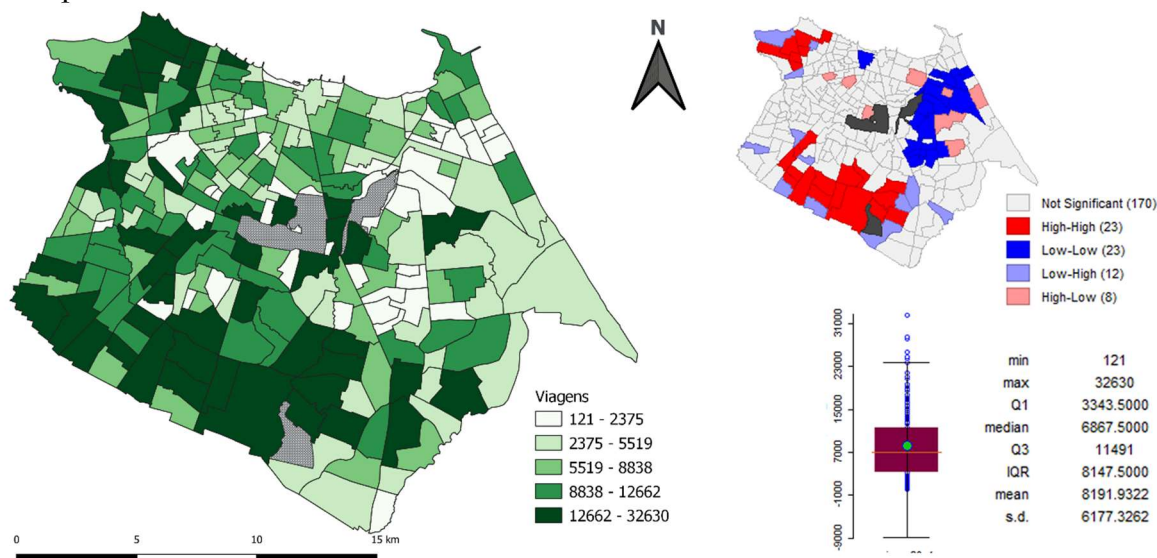


Fonte: Elaborado pelo autor.

A investigação da hipótese que usuários de uso dependente e mais regular está presente em zonas mais periféricas através do indicador de número de dias, só é corroborada pela análise do padrão de usuários com uso dependente e regular (Padrão 3). Para os outros padrões, diferente do esperado, o indicador aumenta nas zonas mais centrais. Este fenômeno pode estar relacionado aos níveis mais altos de acessibilidade dessas zonas, e a não dependência do usuário em relação ao uso do sistema.

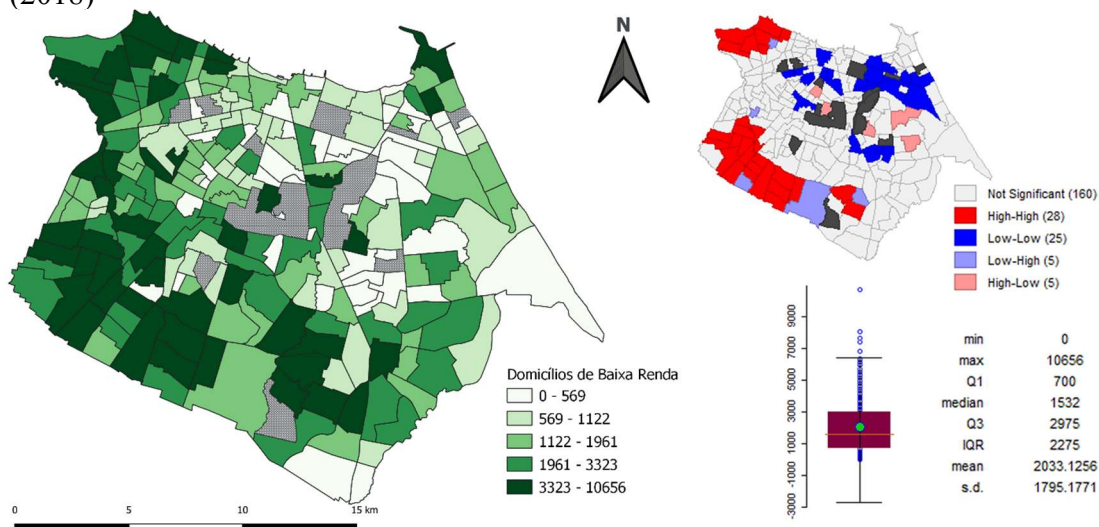
Em seguida, é avaliado o indicador de viagens por domicílios. Esse indicador se propõe a avaliar dois aspectos que impactam na mobilidade de uma zona, a quantidade de viagens produzidas (Figura 31) e a quantidade de domicílios de baixa renda (Figura 32), que são os domicílios responsáveis por gerar a maioria dessas viagens. O intuito é calcular uma taxa de mobilidade para as zonas e identificar a que possuam níveis maiores ou menores dessa taxa de mobilidade, e assim, avaliar a existência de uma desigualdade na distribuição da mobilidade no território. Para isso, inicialmente foram avaliadas as distribuições espaciais das viagens produzidas em um dia útil e dos domicílios de baixa renda. As duas distribuições (Figura 31 e 32) são semelhantes, possuindo maiores valores em zonas mais periféricas e menores valores nas zonas centrais. É possível identificar no Lisa Map das duas variáveis a existência de *clusters High-High* nas regiões sul e noroeste, e *clusters Low-Low* na região central e nordeste.

Figura 31 – Mapa de quintis; Lisa Map e Boxplot para a viagens produzidas em um dia útil por zona no ano de 2018



Fonte: Elaborado pelo autor.

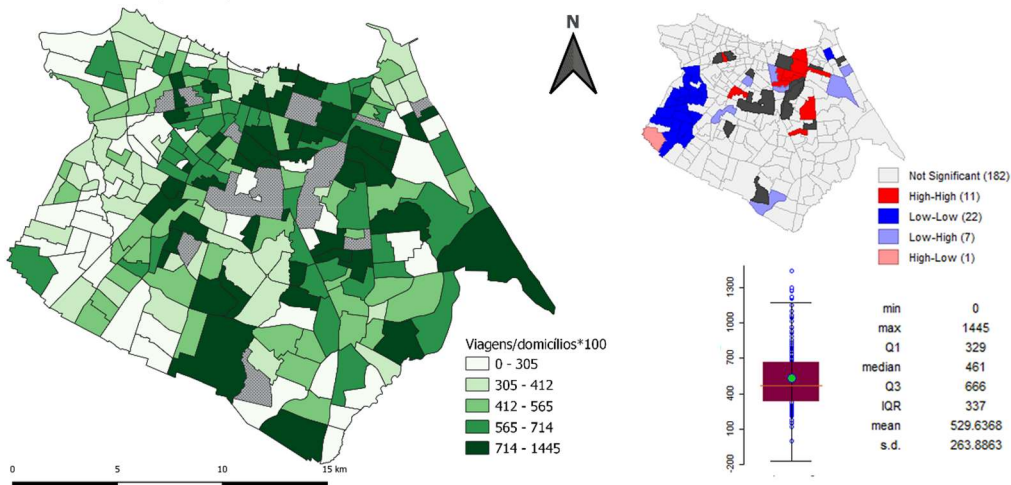
Figura 32 – Mapa de quintis; Lisa Map e Boxplot para os domicílios de baixa renda (2018)



Fonte: Elaborado pelo autor.

Porém, a distribuição do indicador de viagens por domicílios (Figura 33) apresentou uma distribuição oposta ao das variáveis que compõem o indicador, indicando um desequilíbrio nas taxas de mobilidade das zonas. As regiões central, nordeste e sudeste são a que possuem os maiores valores. E as regiões noroeste, sudoeste apresentam os menores valores. Também é possível perceber um aumento do indicador à medida que se distancia das zonas periféricas e se aproxima das zonas mais centrais, com exceção de algumas zonas na região sudeste.

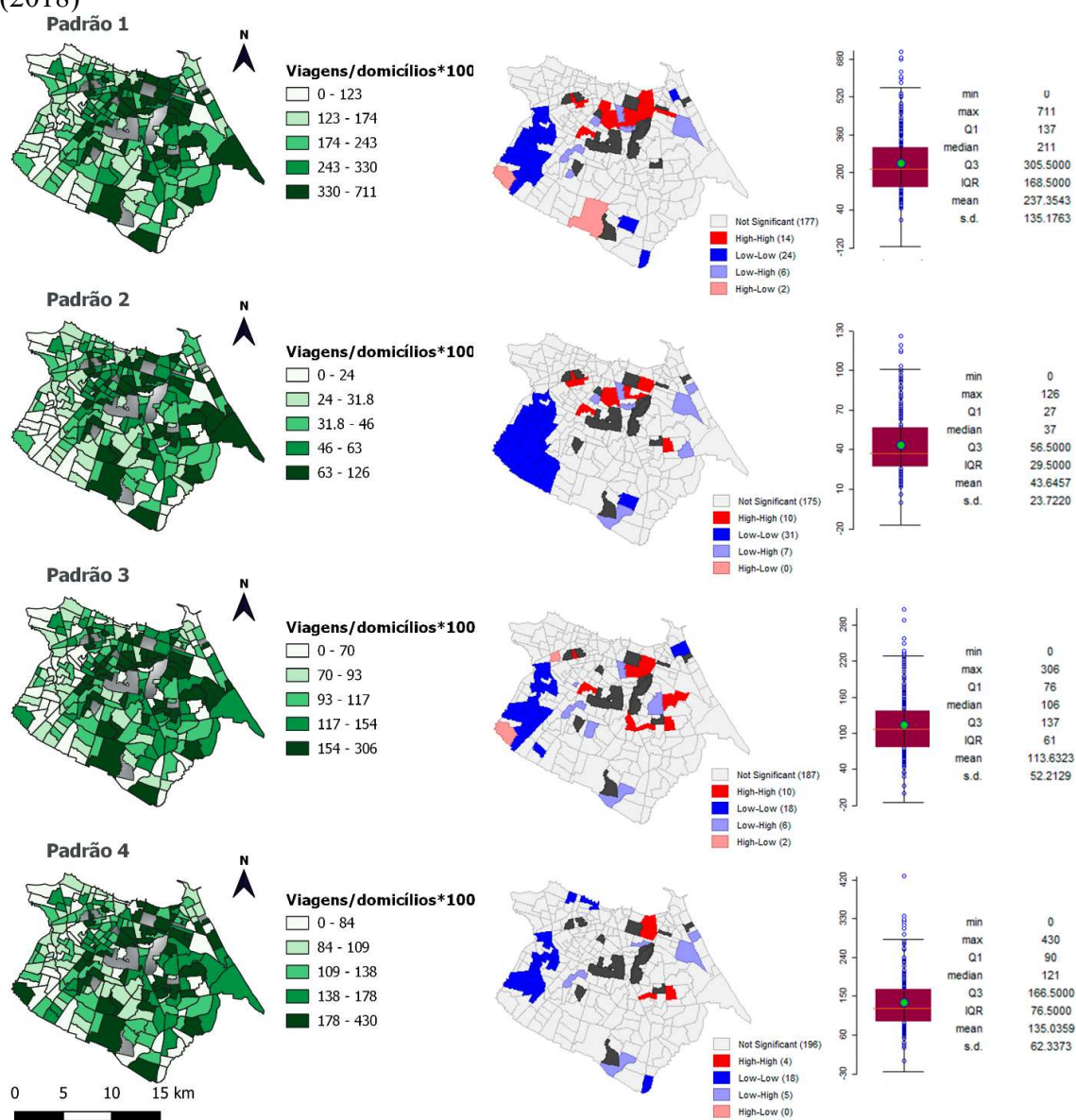
Figura 33 – Mapa de quintis; Lisa Map e Boxplot para o indicador viagens por domicílios (2018)



Fonte: Elaborado pelo autor.

No Lisa Map é possível identificar um *cluster Low-Low* bem definido na região sudoeste e um *cluster High-High* na região centro/nordeste. A partir da apresentação das distribuições espaciais das duas variáveis que compõem o indicador e do próprio indicador, percebe-se que há uma desigualdade na espacialidade da mobilidade no território de Fortaleza, corroborando com a hipótese. As distribuições de cada tipo de padrão (Figura 34) demonstraram ser parecidas ao comportamento geral do indicador.

Figura 34 – Mapa de quintis; Lisa Map e Boxplot para o indicador viagens por domicílios (2018)



Fonte: Elaborado pelo autor.

A desigualdade identificada através da taxa de mobilidade pode ser justificada por diferentes aspectos, como os baixos níveis de acessibilidade das regiões periféricas, por conta elevadas das distâncias entre as residências e as atividades ou pela ineficiência na oferta do sistema. O alto índice de desemprego e de empregos informais, que afeta em maior escala a população de baixa renda, também pode ser fator impactante para essa desigualdade.

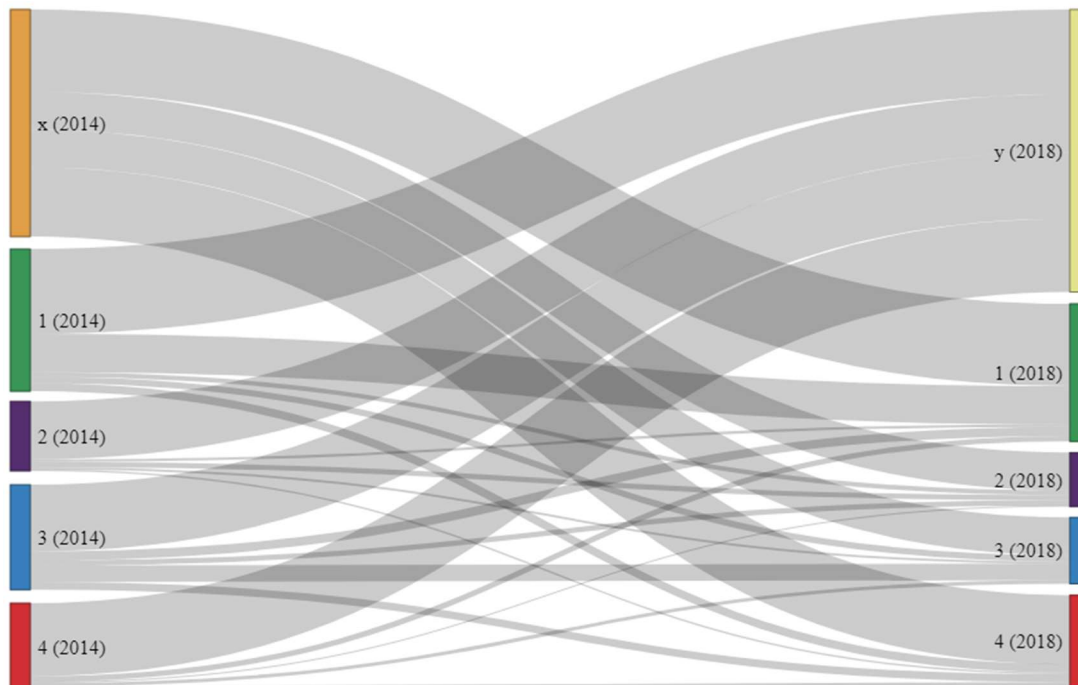
As análises realizadas até aqui permitiram a investigação e corroboram com as hipóteses da existência de diferentes padrões de mobilidade para os usuários de transporte público por ônibus em Fortaleza e de que os padrões mais regulares e dependentes estão mais presentes nas zonas periféricas.

Com objetivo de para facilitar a interpretação das análises seguintes, segue um resumo dos resultados. Os usuários regulares e dependentes do sistema são os usuários do padrão 3, enquanto os usuários esporádicos são os usuários do padrão 1. E os usuários dependentes do sistema, mas não regulares, são os usuários dos padrões 2 e 4. Para esses usuários dependentes, mas regulares (Padrões 2 e 4), as principais diferenças estão na intensidade de uso diário e na distância temporal média. Sendo os usuários do padrão 2 os que apresentaram uma frequência média maior, e a distância temporal média menor quando comparados ao padrão 4. O que talvez possa diferenciar esses dois padrões seja o tipo de atividade realizada. Os resultados sugerem que os usuários do padrão 2 não realizam atividade do tipo trabalho formal, uma vez que a distância temporal média é 5.2 horas, diferentemente dos usuários do padrão 4, que apresenta uma distância temporal média de 7.2 horas.

6.2.3 Análise de evolução temporal

Para avaliar a evolução temporal dos padrões entre os períodos de 2014 a 2018, foi gerado um diagrama de fluxos que permite visualizar como esses padrões se alteram nesse período. Na Figura 35, o grupo x representa os usuários que podem ter começado a utilizar no sistema entre os anos 2015 e 2018, e o grupo y representa os usuários que pararam de utilizar o sistema entre 2014 e 2017. A partir deste gráfico, é possível verificar uma alta taxa de renovação de usuários, com aproximadamente 65% (Grupo x) de novos usuários utilizando o sistema a partir de 2015, e 69.7% (Grupo y) de usuários deixando o sistema até 2017.

Figura 35 – Diagrama de Sankey do fluxo de usuários entre os padrões de 2014 para 2018



Fonte: Elaborado pelo autor.

Além disso, é possível verificar que, dos usuários que permanecem utilizando o sistema, poucos são os que mantêm o tipo de padrão de mobilidade. Sendo o padrão 1, seguido pelo 3, os que mantêm a maior quantidade de usuários no mesmo padrão. Uma maior proporção de usuários que deixam o sistema é observada nos Grupos 1 e 4, possivelmente usuários do tipo não-cativos que migraram para outro modo de transportes. Os usuários dos padrões 1 podem estar mais suscetíveis a alguma mudança de posse de veículo ao longo dos anos devido à grande parcela de estudantes neste grupo.

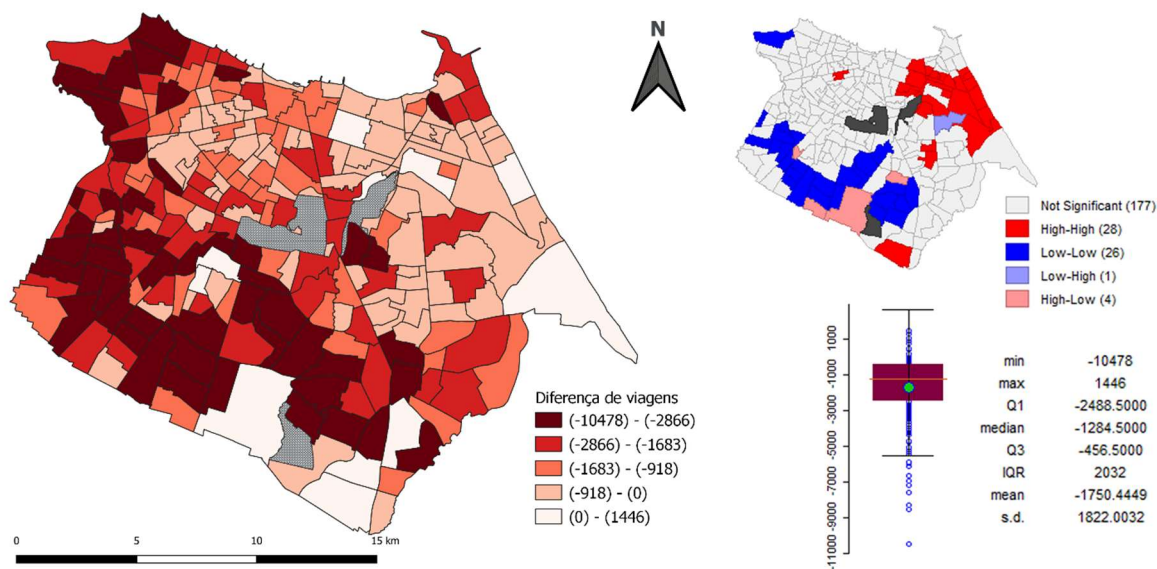
Apesar do Padrão 1 ser responsável pela maior quantidade de usuários deixando o sistema, também é o padrão que recebe mais usuários no ano de 2018. O padrão de usuários dependentes e regulares (Padrão 3) é o que mais reduz o seu tamanho entre o período analisado, ou seja, perdeu mais usuários do que ganhou. Isso indica que este padrão é o mais afetado pela queda na demanda, uma vez que é o padrão responsável por produzir mais viagens, já que os seus usuários apresentam uma frequência média maior e um uso do sistema ao longo do ano mais regular.

Alguns questionamentos surgem a partir dos resultados dessas análises, como: O fenômeno de renovação dos usuários para um período de cinco anos é este mesmo? Poderia, parte dos usuários, estarem trocando de cartão, e conseqüentemente de código de identificação, durante o período de análise? Houve migração dos usuários para outro modo de transporte? Alguma mudança nas características socioeconômicas influenciou essa renovação? Houve mudança no endereço dos usuários por conta do espraiamento urbano? Quais fatores influenciam para uma taxa de renovação tão elevada? São questionamento que precisam ser investigados em análises futuras.

6.2.4 Análise de evolução espacial

Para avaliar a evolução espacial desses padrões foram analisados os indicadores: Diferença de viagens por domicílios de baixa renda, diferença de viagens produzidas, diferença de usuários. As distribuições espaciais dos indicadores de viagens produzidas e de usuários são bem parecidas, por isso serão apresentados a seguir somente os indicadores de diferença de viagem por domicílios de baixa renda e diferença de viagens. As distribuições do indicador de diferença de usuários podem ser encontradas no apêndice. Os indicadores foram analisados de modo geral (Figura 36) e para cada padrão individualmente (Figura 37).

Figura 36 – Mapa de quintis; Lisa Map e Boxplot para o indicador diferença de viagens entre 2018 e 2014



Fonte: Elaborado pelo autor.

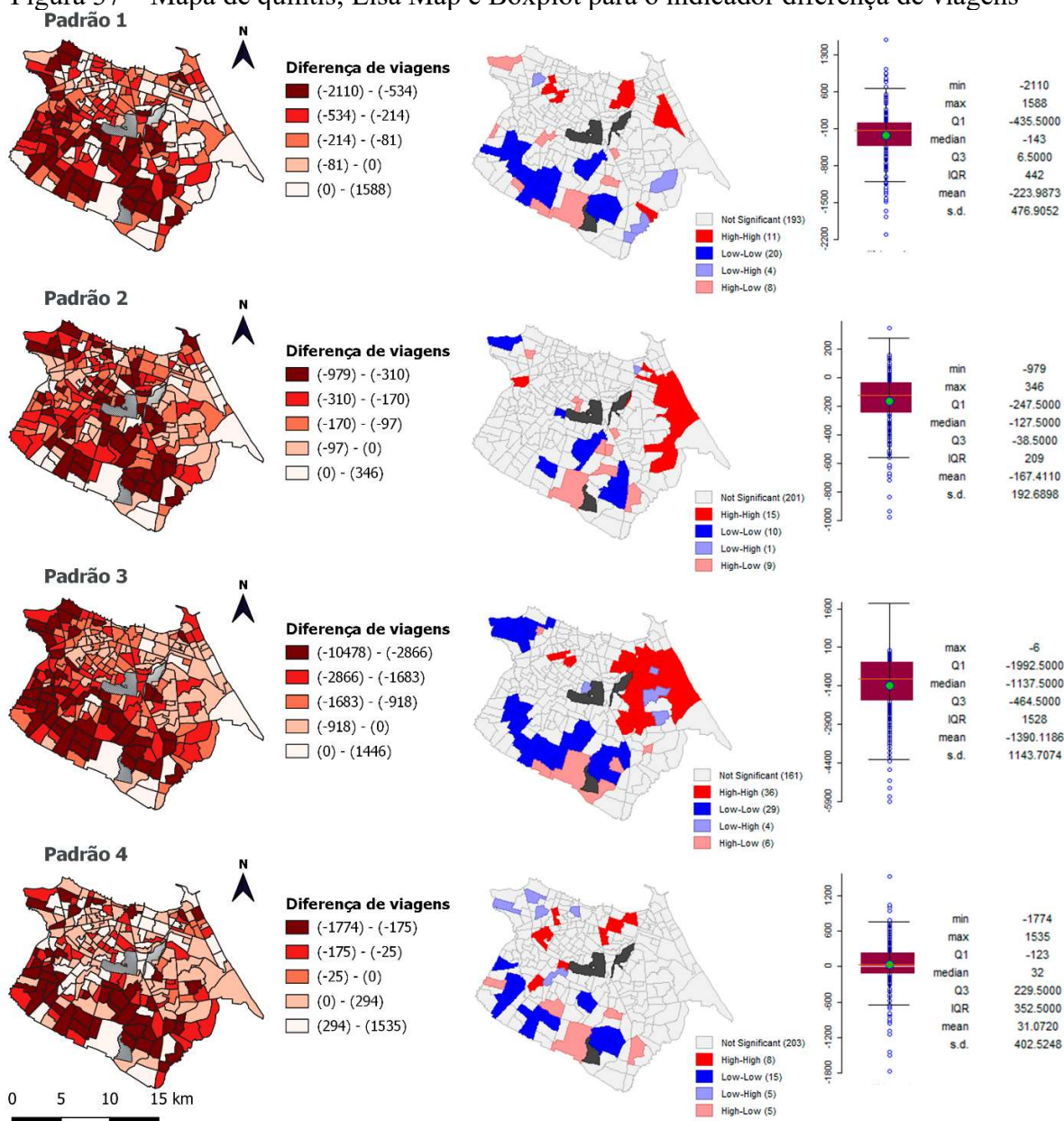
A distribuição espacial do indicador de diferença de viagens para todos os padrões apresentou que no geral as quedas na produção de viagens ocorrem de forma bem mais intensa nas zonas periféricas e diminui à medida que se aproxima das regiões centrais. Algumas zonas categorizadas como *High-Low* foram identificadas nas zonas da região sul e sudeste, que podem ser resultado do fenômeno do espraiamento urbano explicado por Lima (2017) e de programas habitacionais como o Residencial Cidade Jardim II realizado em conjunto pela Prefeitura de Fortaleza e o Governo do Estado do Ceará e lançado no ano de 2018.

Conforme esperado, na Figura 37 é possível visualizar que o padrão dos usuários regulares e dependentes do sistema (Padrão 3) é o mais afetado pelo o problema de queda na demanda. No geral, as distribuições espaciais de todos os padrões são parecidas, porém algumas observações merecem destaques.

Nos padrões de usuários esporádicos (Padrão 1) e não regulares, porém dependentes que realizam atividades trabalhos (Padrão 4), algumas zonas pontuais mais próximas ao centro da cidade e aos corredores centrais possuem um aumento na quantidade de viagens. Ou seja, existe um aumento de usuários não dependente do sistema, e uma queda na quantidade de usuários dependentes, o que sugere um aumento da distribuição desigual da mobilidade. Isto evidência que existe uma distribuição insustentável da mobilidade para os usuários de transporte público, uma vez que o problema da desigualdade apresenta um agravamento ao longo do tempo. Este fenômeno pode estar acontecendo por alguma mudança na oferta no sistema ou pela possibilidade dos usuários de integrarem com outros modos.

Os padrões de usuários não regulares, porém dependentes que realizam atividades do tipo outros ou educação (Padrão 2) e de usuários regulares e dependentes (Padrão 3) apresentaram um *cluster High-High* nas zonas da região nordeste, mas o que o mapa de quintis indica é que apenas a queda na demanda não foi tão intensa nessa região.

Figura 37 – Mapa de quintis; Lisa Map e Boxplot para o indicador diferença de viagens

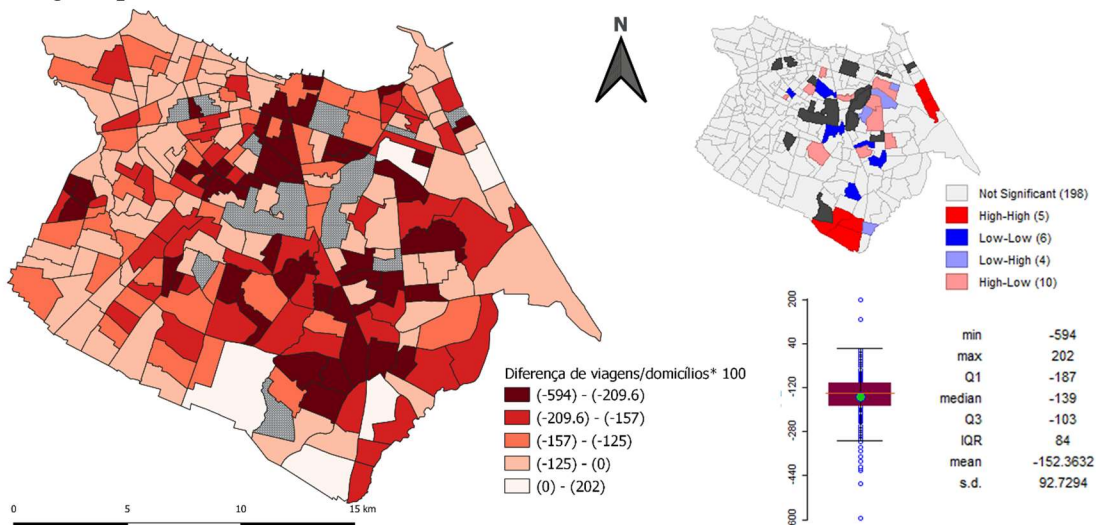


Fonte: Elaborado pelo autor

Em seguida, foi avaliado o indicador de diferença de viagens por domicílios (Figura 38). Este indicador apresentou uma queda para quase toda as zonas, com poucas exceções. Algumas zonas centrais e na região sudeste apresentaram uma queda mais intensa, o que mostra, que o problema de distribuição desigual pode estar reduzindo. Os domicílios utilizados para o ano de 2018 são apenas valores expandidos para os valores utilizados no ano de 2014. Logo, as

diferenças desse indicador são causadas, basicamente, pela redução nas viagens produzidas por zona.

Figura 38 – Mapa de quintis; Lisa Map e Boxplot para o indicador diferença de viagens por domicílios



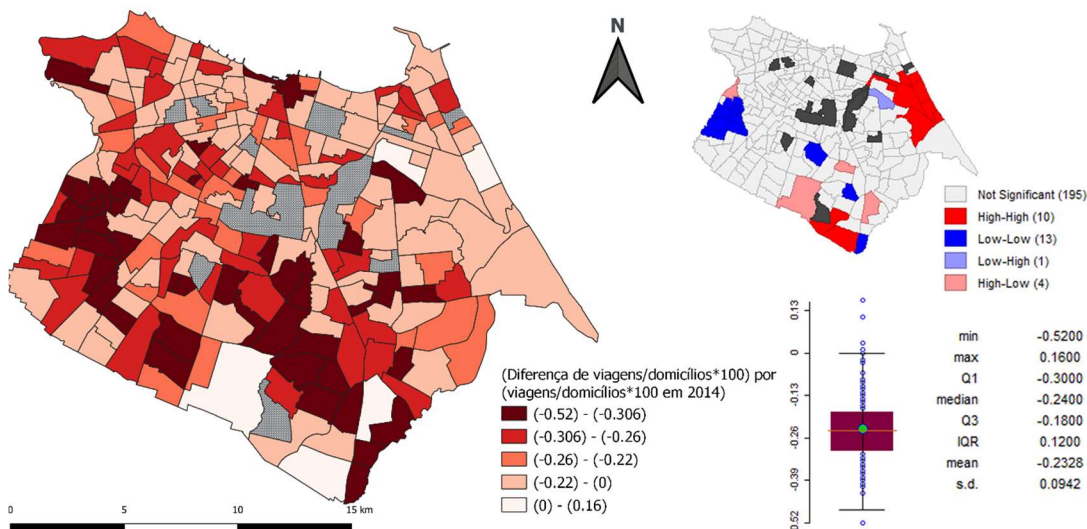
Fonte: Elaborado pelo autor.

Para avaliar se essas quedas foram relevantes ao valor apresentado no ano de 2014, foi avaliado a proporção entre a diferença de viagens por domicílios pelo o valor do indicador de viagens de domicílios em 2014 (Figura 39). Esse indicador apresentou uma situação diferente do indicador de diferença de viagens por domicílios (Figura 38). O Lisa Map deste último indicador apresentou um *cluster Low-Low* na região sudoeste, indicando que essa região possuem uma concentração de zonas que apresentaram queda no indicador de viagens por domicílios. É possível também visualizar um *cluster High-High* na região nordeste, indicando que essas zonas pouco sofreram alterações, ou que possuíram um leve aumento do indicador.

As zonas centrais que apresentavam uma alta queda na Figura 38, não apresentam uma a mesma intensidade para o indicador de proporção (Figura 39). E algumas zonas nas regiões sudoeste e sudeste apresentaram uma queda proporcionalmente maior ao seu valor efetivo de 2014. Ou seja, por mais que o valor bruto da queda na taxa de mobilidade tenha ocorrido em algumas zonas centrais de forma mais intensa, indicando uma redução na desigualdade, o valor relativo da queda em relação aos valores de 2014 indicam que a taxa de mobilidade sofre uma redução

proporcionalmente maior nas zonas periféricas. O que indica que na verdade essa desigualdade não está diminuindo, e sim piorando.

Figura 39 – Mapa de quintis; Lisa Map e Boxplot para o indicador proporção da diferença das viagens por domicílios pelo o valor das viagens por domicílios no ano de 2014

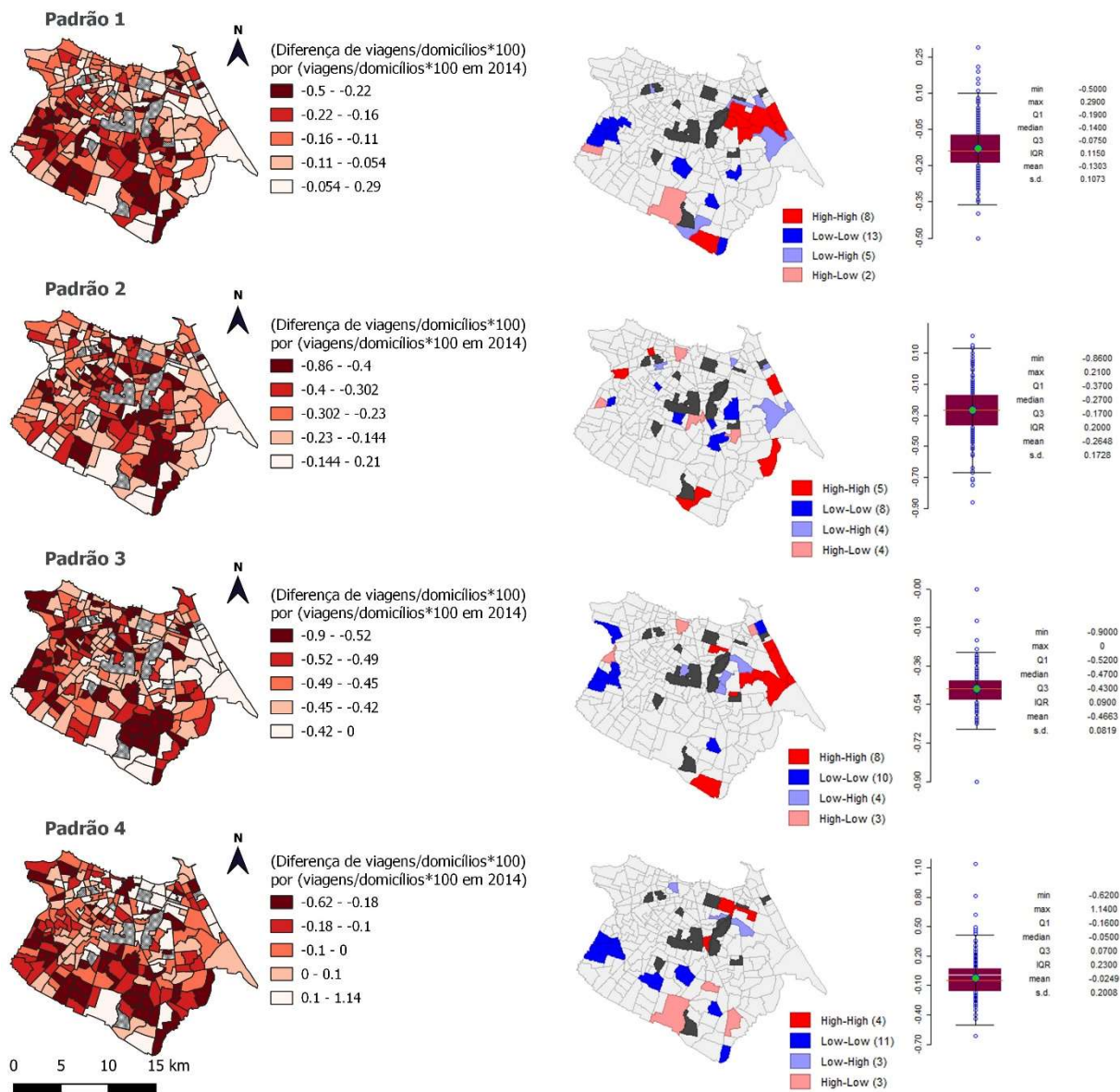


Fonte: Elaborado pelo autor.

A distribuição do indicador para cada padrão (Figura 40) apresentou resultados parecidos com a distribuição geral dos usuários. Mas vale destacar que o padrão de usuários não regulares, porém dependentes que realizam atividades do tipo outros ou educação (Padrão 2), que não apresentou valores tão intensos nas regiões periféricas. E o padrão de usuários dependentes e regulares (Padrão 3) que também apresentou quedas proporcionalmente significativa nas regiões periféricas.

Os resultados dessas análises reforçam as hipóteses de que usuários de padrões de uso dependente e mais regular são mais impactados pela queda na demanda que padrões menos regulares são menos impactados. Os resultados mostram também que a taxa de mobilidade para algumas zonas da periferia está sendo proporcionalmente mais impactada pela queda na demanda do que as zonas centrais que apresentam uma queda bruta maior. Sabe-se que existe uma distribuição desigual da mobilidade para os usuários de transporte público, e que essa distribuição desigual está piorando ao longo dos anos, o que indica também um problema de distribuição insustentável da mobilidade para os usuários dependentes desse sistema.

Figura 40 – Mapa de quintis; Lisa Map e Boxplot para o indicador proporção da diferença das viagens por domicílios pelo o valor das viagens por domicílios de 2014 para cada padrão



Fonte: Elaborado pelo autor.

7 CONCLUSÕES E RECOMENDAÇÕES

7.1 Conclusões

Este trabalho teve como objetivo geral caracterizar os padrões de mobilidade dos usuários de transporte público por ônibus de Fortaleza através da utilização de *Big Data* das validações gerado pelo sistema de *Smart Card*. O caráter contínuo e quase populacional desses dados, diferentemente das bases de dados comumente utilizadas no planejamento de transportes, permitiu a geração de atributos e indicadores que permitem a identificação de características da mobilidade que não é possível através de pesquisas domiciliares.

De forma geral, a revisão de literatura realizada mostrou que a demanda, ou mobilidade da demanda é comumente caracterizada de modo a considerar que esta possui comportamentos homogêneos para um determinado nível de agregação por zona. Porém, uma revisão de literatura mais conceitual mostrou que as características que impactam na realização de uma viagem mostram que é improvável que os comportamentos de viagem de usuários sejam os mesmos para uma mesma zona, ou para um mesmo grupo social. Esse trabalho se propôs a superar essa limitação através da aplicação de técnicas de clusterização para combinar atributos que representem a mobilidade dos usuários e, assim, identificar padrões de mobilidade dos usuários de transporte público por ônibus em Fortaleza.

Além disso, foi caráter motivacional desse estudo a queda na demanda do sistema de transporte público por ônibus, que, a priori, é um problema para as empresas operadoras dos sistemas e para os planejadores da cidade, mas também pode ser um problema para a sociedade em geral. Por isso, esse estudo considerou que compreender o comportamento desses usuários, conhecendo melhor as características dos usuários que estão deixando de utilizar o sistema é tarefa importante para garantir a sustentabilidade do sistema de transporte público por ônibus e entender melhor as dinâmicas sociais que envolvem deslocamentos na cidade. Assim, considera-se que a aplicação do método proposto, pode fornecer informações importantes para os gestores e operadores sobre os diferentes segmentos de usuários, e como estes segmentos vêm se alterando ao longo do tempo.

O método de clusterização utilizado apresentou resultados que permitiram evidenciar a existência de diferentes grupos de mobilidade para os usuários de transporte público por ônibus

de Fortaleza. A partir de uma revisão de literatura, de um método estatístico e uma análise descritiva dos dados, foram definidos quatro grupos de usuário para representar os padrões de mobilidade. Esses padrões foram interpretados de acordo com a dependência do sistema de transporte coletivo, a sua regularidade e o tipo de atividade (trabalho formal, trabalho informal, estudo e outros). O estudo mostrou que os usuários classificados como de uso esporádico (não-cativos e que usam o sistema para acessar atividades por motivos outros) representam a maior parcela de usuários, 37% dos usuários entre os anos 2014 e 2018. Enquanto os usuários de uso dependente e regular (cativos e que usam o sistema para acessar atividade de trabalho) e que, portanto, mais contribuem para a receita do sistema, representam uma parcela de 23% dos usuários.

As análises de variação temporal ao longo dos cinco anos evidenciaram que a quantidade de usuários que deixaram o sistema no período analisado, em torno de 70%, é maior do que a quantidade de usuários ingressantes, em torno de 65%. Os usuários que mais deixaram o sistema, provocando uma queda expressiva no número de validações, foram os usuários cativos que possivelmente usam o sistema para acessar atividade de trabalho. As variações mostraram também que, dos usuários que continuam utilizando o sistema durante este período, poucos são os que mantêm o padrão de mobilidade.

As análises dos padrões espaciais em cada ano reforçaram a hipótese de que os usuários que dependem do sistema para acessar atividades de trabalho estão localizados nas zonas periféricas de Fortaleza, distantes das zonas de emprego. As análises também apoiam a ideia de que usuários não-cativos que usam o sistema com outros propósitos, que não trabalho, estão nas zonas com melhores níveis de acessibilidade da cidade. Os resultados evidenciam, assim, problemas de distribuição desigual da mobilidade em que os usuários que mais precisam do sistema estão mais distantes das oportunidades de emprego. Além disso, revelam a tendência de uso esporádico ou irregular em regiões centrais onde possui melhores níveis de acessibilidade.

A análise de variação destes padrões espaciais entre os anos 2014 e 2018 mostrou uma queda expressiva de demanda em zonas periféricas da cidade de Fortaleza, especialmente para o grupo de usuários classificados como cativos e que usam o sistema para acessar atividades de trabalho. Este resultado reforça a hipótese de que os problemas de mobilidade do sistema de transporte público vêm se agravando ao longo dos anos, especialmente para os usuários localizados nas zonas periféricas e que, provavelmente, pertencem ao grupo socialmente mais vulnerável. O que demonstra também a existência de problemas de uma distribuição insustentável

da mobilidade. A queda da demanda foi percebida em todos os padrões e em quase toda as zonas, porém, foi percebido um aumento na demanda para os padrões de usuários não cativos, em algumas zonas centrais específicas que estão próximas aos corredores centrais da cidade. Este fenômeno pode estar acontecendo por conta de uma melhora no nível dos serviços, e pela possibilidade de integração com outros modos de transporte.

Possivelmente, as mudanças dos padrões e renovação de usuário podem ser explicadas pelo fenômeno de espraiamento urbano e, portanto, do desequilíbrio na distribuição do uso solo, como discutido em Lima (2017) Outros fatores, como, novos modos de transporte (uber, posse de automóvel, bicicleta compartilhadas) e/ou aumento no índice de desemprego da população também podem ser fatores que estejam influenciando nessas dinâmicas.

Por tanto, considera-se que o método aplicado contribuiu para identificação e caracterização dos padrões de mobilidade dos usuários de transporte público por ônibus de Fortaleza. De certa forma, os atributos e método de clusterização escolhidos foram capazes de perceber e diferenciar alguma heterogeneidade na demanda. A avaliação da queda da demanda a partir dos padrões identificados permitiu identificar quais são os usuários que são mais impactados por esta queda.

7.2 Limitações e recomendações

A partir do método proposto algumas limitações podem ser identificadas. Primeiramente, limitações em relação aos dados. Parte considerável dos usuários não permitiram a identificação do endereço de origem através do endereço de residência buscados nas bases de cadastros, por tanto, sugere-se que as análises sejam refeitas quando uma base de cadastro mais completa estiver disponível. Vale ressaltar que este estudo, em algumas de suas análises, não considerou as viagens do tipo Inteira (à dinheiro) que contribuem com uma parcela expressiva da receita do sistema. Porém, acredita-se que essa limitação será amenizada nos próximos anos, pois as empresas que operam o sistema estão removendo gradativamente a opção de pagamento por dinheiro, fazendo com que os usuários adotem ao pagamento por cartão, permitindo assim rastrear a mobilidade desses usuários.

Os dados de domicílios também foram, de certo modo, fator limitante para análises do indicador de viagens por domicílios uma vez que algumas zonas tiveram que ser desconsideradas.

Por tanto, recomenda-se que a análise do indicador seja refeita quando houver disponíveis dados mais recentes e completos dos domicílios.

Apesar da análise realizada para averiguar as distâncias entre a parada mais próxima da residência que pertence a linha mais utilizada do usuário no período da manhã e o endereço de residência dos usuários de cadastros, esse método de estimar a origem da viagem também pode ser fator limitante. Sugere-se então, a partir dos dados de *Smart Card*, a construção de um modelo de modelo probabilístico para estimar a origem da viagem com base em atributo que possam influenciando o local de validação, como: a linha do embarque, o número de domicílios por zona, a influência dos terminais físicos e a distância em relação a localização do centroide de validações. Com o modelo construído os resultados dos dois métodos podem ser comparados e então avaliados.

Outro aspecto que é caráter limitante do método proposto é em relação aos atributos selecionados para identificar os padrões. Conforme visto na literatura, as decisões de viagens são influenciadas por diferentes aspectos. A consideração de mais aspectos, como outras características socioeconômicas, como renda, posse de veículo, atividade principal realizada, entre outras, podem ajudar no processo de identificação e interpretação dos padrões de mobilidade. Como, por exemplo, na interpretação dos padrões dependentes, mas não regulares (Padrões 2 e 4). Acredita-se que se houvesse algum atributo que remetesse de forma mais direta a atividade desses usuários, estes padrões pudessem ser mais facilmente interpretados.

Ainda em relação aos dados, é importante ressaltar que a característica contínua desses dados pode contribuir para a realização de análises de avaliação *ex-post* de intervenções realizadas no sistema. Permitindo assim, avaliar como os padrões de mobilidade dos usuários se modificaram após a realização de alguma intervenção. Portanto, recomenda-se a utilização deste método de identificação de padrões em combinação com algum método de avaliação *ex-post* de intervenções nos sistemas de transportes, a fim de compreender se essas intervenções têm alcançado os objetivos esperados.

Para as análises temporais, optou-se por analisar o retrato dos períodos de 2014 e 2018. Porém, os resultados apresentaram um fenômeno interessante e inesperado de renovação de usuários durante o período. Para entender melhor esse fenômeno, assim como ajudar a levantar e checar hipóteses sobre ele, sugere-se uma análise ano a ano, combinada a uma contextualização das mudanças que aconteceram na oferta do sistema. Deste modo, seria possível observar mais detalhes sobre as mudanças dos padrões de mobilidade.

Acredita-se que alguns fatores principais afetaram a queda na demanda do sistema de transporte público por ônibus em Fortaleza, como: Novos modos disponíveis; Desemprego; Baixos níveis de acessibilidade e Mudanças no uso do solo. Apesar das análises realizadas para todos os padrões, acredita-se que para entender melhor o comportamento dos usuários que deixaram de utilizar o sistema, a realização de análises descritivas e espaciais específicas para esses usuários, pode ajudar na realização de uma melhor caracterização destes, assim como uma identificação de potenciais motivos que levaram estes usuários a deixarem de utilizar o sistema.

É importante citar que as análises realizadas nesse estudo têm objetivo apenas de caracterizar os padrões de mobilidade identificados, não sendo objetivo realizar nenhum tipo de diagnóstico ou identificar relações de causalidade entre as variáveis e os problemas identificados. Para isto, recomenda-se a realização de estudo específico que utilize de técnicas mais eficazes para identificar e mensurar causalidades, como, por exemplo, as técnicas de modelagem de equações estruturais.

REFERÊNCIAS

AGARD, B.; MORENCY, C.; TRÉPANIÉ, M. Mining Public Transport User Behaviour From Smart Card Data. **IFAC Proceedings Volumes**, v. 39, n. 3, p. 399-404, 2006.

AGARD, B.; PARTOVINIA, V.; TREPANIÉ, M. Assessing public transport travel behaviour from smart card data with advanced data mining techniques. *In: WCTR – WORLD CONFERENCE ON TRANSPORTATION RESEARCH*, 13., 2013, Rio de Janeiro. **Proceedings** [...]. Rio de Janeiro, 2013.

AGHABOZORGI, S.; SEYED SHIRKHORSHIDI, A.; YING WAH, T. Time-series clustering– A decade review. **Inf Syst.**; v. 53, p. 16-38, 2015.

AKMAN, O.; COMAR, T.; HROZENCIK, D.; GONZALES, J. Data Clustering and Self-Organizing Maps in Biology. **Algebraic and Combinatorial Computational Biology**, p. 351-374, 2019.

ALI, A.; KIM, J.; LEE, S. Travel Behavior Analysis Using Smart Card Data. **KSCE J. Civ. Eng.**, v. 20, p. 1532-1539, 2016.

ANDA, C.; FOURIE, P.; ERATH, A. Transport Modelling in the Age of *Big Data*. *Journal of Urban Sciences*, v. 21, n. S1, p. 19-42, 2017. Disponível em: <https://www.research-collection.ethz.ch/handle/20.500.11850/118462>. Acesso em: 17 de nov. de 2020.

ANDRADE, B. R. **Compreensão da problemática da periferização por segregação involuntária no planejamento da acessibilidade e mobilidade em Fortaleza**. 2016. Dissertação (Mestrado em Engenharia de Transportes) – Universidade Federal do Ceará, Fortaleza, 2016.

AUDOUIN, M.; RAZAGHI, M.; FINGER, M. How Seoul Used the ‘t-money’ Smart Transportation Card to Re-plan the Public Transportation System of the City, Implications for Governance of Innovation in Urban Public Transportation Systems. *In: TRANSIST SYMPOSIUM*, 8., 17-18 dez. 2015, Istanbul, **Proceedings** [...]. Istanbul, 2015.

BALCOMBE, R.; MACKETT, R.; PAULLEY, N.; PRESTON, SHIRES, J.; TITHERIDGE, H.; WARDMAN, M.; WHITE, P. The demand for public transport: a practical guide. **TRL Reports. Report TRL**, v. 593, p. 1-238, 2004.

BAGCHI, M.; WHITE, P. R. The potential of public transport smart card data. **Transport Policy**, v. 12, n. 5, p. 464-474, 2005.

BANISTER, D. The sustainable mobility paradigm. **Transport Policy**, v. 15, p. 73-80, 2008.

BARROSO, L. P.; ARTES, R. **Análise multivariada**. Lavras: UFLA, 2003. 151 p.

BEDER, S. Costing the Earth: Equity, Sustainable Development and Environmental Economics. **N Z J Environ Law**, v. 4, p. 227-243, 2000.

BERTHOLD, M. R.; HÖPPNER, F. **On Clustering Time Series Using Euclidean Distance and Pearson Correlation**. ArXiv: 1601.02213v1 [cs.LG], 20 jan. 2016. Disponível em: <https://arxiv.org/pdf/1601.02213.pdf>. Acesso em: 20 de nov. de 2020.

BHAGAT, A.; KSHIRSAGAR, N.; KHODKE, P.; DONGRE, K.; ALI, S. Penalty Parameter Selection for Hierarchical Data Stream Clustering. **Procedia Computer Science**, v. 79, p. 24 - 31, 2016.

BLYTHE, P.; BRYAN, H. Understanding Behaviour through Smartcard Data Analysis. **Transport**, v. 160, p. 173-177, 2007.

BOSCHMANN, E. E.; KWAN, M.-P. Toward Socially Sustainable Urban Transportation: Progress and Potentials. **International Journal of Sustainable Transportation**, v. 2, n. 3, p. 138-157, 2008.

BRAGA, C. **Big Data de Transporte Público na Análise da Variabilidade de Indicadores da Acessibilidade às Oportunidades de Trabalho e Educação**. 2019. Dissertação (Mestrado em Engenharia de Transportes) – Universidade Federal do Ceará, Fortaleza, 2019.

BRIAND, A. S.; CÔME, E.; TRÉPANIER, M.; OUKHELLOU, L. Analyzing Year-to-year Changes in Public Transport Passenger Behaviour Using Smart Card Data. **Transp. Res. Part C Emerg. Technol.**, v. 79, p. 274-289, 2017.

CARDOSO, C. E. P. **Análise do transporte coletivo urbano sob a ótica dos riscos e carências sociais**. 2008. Tese (Doutorado) – Programa de Pós-Graduação em Serviço Social, Pontifícia Universidade Católica de São Paulo, São Paulo, 2008.

CHU, K. K.; CHAPLEAU, R. (2007). Imputation Techniques for Missing Fields and Implausible Values in Public Transit Smart Card Data. *In: WORLD CONFERENCE ON TRANSPORTATION RESEARCH*, 11., 2007, Berkeley. **Proceedings [...]**. Berkeley, 2007.

COSTA, M. S. **Um índice de mobilidade urbana sustentável**. 2008. Tese (Doutorado em Planejamento e Operação de Sistemas de Transportes) – Universidade de São Paulo, São Paulo, 2008.

COSTA, P. B.; NETO, G. C. M.; BERTOLDE, A. I. Urban Mobility Indexes: A Brief Review of the Literature. **Transportation Research Procedia**, v. 25, p. 3645-3655, 2017.

DEVILLAIN, F.; MUNIZAGA, M.; TRÉPANIER, M. Detection of Activities of Public Transport Users by Analyzing Smart Card Data. **Transportation Research Record: Journal of the Transportation Research Board**, v. 2276, n. 1, p. 48-55, 2013.

EL MAHRISI, M. K.; CÔME, E.; BARO, J.; OUKHELLOU, L. (2014). Understanding passenger patterns in public transit through Smart Card and socioeconomic data. *In:*

INTERNATIONAL WORKSHOP ON URBAN COMPUTING (URBCOMP), 3., ACM SIGKDD CONFERENCE, 2014, New York. **Proceedings** [...]. New York, USA, 2014.

EL MAHRISI, M. K.; COME, E.; OUKHELLOU, L.; VERLEYSSEN, M. Clustering Smart Card Data for Urban Mobility Analysis. **IEEE Transactions on Intelligent Transportation Systems**, v. 18, n. 3, p. 712-728, 2017.

ESTER, M.; KRIEGEL, H-P.; SANDER, J.; XU, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *In*: INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 2., 1996. **Proceedings** [...].1996. p. 226-231.

FAROQI, H.; MESBAH, M.; KIM, J. Comparing Sequential with Combined Spatiotemporal Clustering of Passenger Trips in the Public Transit Network Using Smart Card Data. **Mathematical Problems in Engineering**, p. 1-16, 2019.

FEITELSON, E. Introducing environmental equity dimensions into the sustainable transport discourse: issues and pitfalls. **Transportation Research Part D: Transport and Environment**, v. 7, n. 2, p. 99-118, 2002.

FORTALEZA. **Brazil Wins 2019 Sustainable Transport Award**. Disponível em: <https://www.itdp.org/2018/06/29/fortaleza-brazil-wins-2019-sta/>. Acesso em: 14 dez. 2019.

FORTALEZA. PREFEITURA MUNICIPAL. **Fortaleza 2040. Cidade Conectada, Acessível e Justa. Plano de Mobilidade e Acessibilidade Urbano**. Fortaleza, 2015. v. 3, partes I, II e III.

FREITAS, A. T. **Metodologia de caracterização da problemática do sistema de transporte público de passageiros a partir dos dados da Bilhetagem Eletrônica**. 2015. Dissertação (Mestrado em Engenharia de Transportes) – Universidade Federal do Ceará, Fortaleza, 2015.

FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R. The elements of statistical learning. **Springer Series in Statistics**, New York, v. 1, 2001.

GARCIA, C. **Strategic Assessment of Accessibility on Urban Mobility Networks**. 2016. Dissertação (Doutorado em Sistemas de Transportes) – Universidade de Lisboa, [s.l.], 2016.

GARCIA, C. S. H. F.; MACÁRIO, M. do R. M. R.; MENEZES, E. D. de A. G.; LOUREIRO, C. F. G. Strategic Assessment of Lisbon's Accessibility and Mobility Problems from an Equity Perspective. **Networks and Spatial Economics**, v. 18, n. 2, p. 415-439, 2018.

GILLIS, D.; SEMANJSKI, I.; LAUWERS, D. How to Monitor Sustainable Mobility in Cities? Literature Review in the Frame of Creating a Set of Sustainable Mobility Indicators. **Sustainability**, v. 8, n. 1, p. 29, 2015.

GSCHWENDER, A.; MUNIZAGA, M.; SIMONETTI, C. (2016). Using *Smart Card* and GPS Data for Policy and Planning: The Case of Transantiago. **Res. Transp. Econ.**, v. 59, p. 242-249, 2016.

HALVORSEN, A.; KOUTSOPOULOS, H. N.; MA, Z.; ZHAO, J. Demand management of congested public transport systems: a conceptual framework and application using Smart Card data. **Transportation**, 2019.

HANDY, S. L. **Accessibility Vs. Mobility**: enhancing Strategies for Addressing Automobile Dependence in the U.S. Davis: Institute of Transportation Studies, University of California, 2002.

HANDY, S. L. **Planning for Accessibility**: in theory and in practice. Access to Destination, Elsevier, 2005.

HATZICHRISTOS, T. Delineation of demographic regions with GIS and computational intelligence. **Environment and Planning B: Planning and Design**, v. 31, p. 39-49, 2004.

HAWKINS, J.; HABIB, K. N. Heterogeneity in marginal value of urban mobility: evidence from a large-scale household travel survey in the Greater Toronto and Hamilton Area. **Transportation**, v. 47, p. 3091-3108, 2020.

HENRIQUE, C. S **Diagnóstico espacial da mobilidade e da acessibilidade dos usuários do sistema integrado de transporte de Fortaleza**. 2004. Dissertação (Mestrado) – Programa de Pós-Graduação em Engenharia de Transportes, Universidade Federal do Ceará, Fortaleza, 2004.

HICKMAN, R.; HALL, P.; BANISTER, D. Planning more for sustainable mobility. **Journal of Transport Geography**, v. 33, p. 210-219, 2013.

IACONO, M.; LEVINSON, D.; EL-GENEIDY, A. Models of transportation and land use change: A guide to the territory. **Journal of Planning Literature**, v. 22, n. 4, p. 323-340, 2008.

IPEA. **Transporte urbano e inclusão social: elementos para políticas públicas**. 2003.

IRANI, J.; PISE, N.; PHATAK, M. Clustering Techniques and the Similarity Measures used in Clustering: A Survey. **International Journal of Computer Applications**, v. 134, n. 7, p. 9-14, 2016.

IWASAKI, Y.; KUSNE, A. G.; TAKEUCHI, I. Comparison of dissimilarity measures for *cluster* analysis of X-ray diffraction data from combinatorial libraries. **Npj Computational Materials**, v. 3, n. 1, 2017.

JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data *clustering*: a review. **ACM Computing Surveys**, v. 31, n. 3, p. 264-323, 1999.

JIN, K. E.; JI, Y. S.; MOON, D. S. Analysis of Public Transit Service Performance Using Transit Smart Card Data in Seoul. **KSCE J. Civ. Eng.**, v. 19, p. 1530-1537, 2015.

KENWORTHY, J. Urban Planning and Transport Paradigm Shifts for Cities of the Post-Petroleum Age. **Journal of Urban Technology**, v. 14, n. 2, p. 47-70, 2007.

KIEU, L. M.; BHASKAR, A.; CHUNG, E. A modified Density-Based Scanning Algorithm with Noise for spatial travel pattern analysis from Smart Card AFC data. **Transportation Research Part C: Emerging Technologies**, v. 58, p. 193-207, 2015.

KIEU, L. M.; BHASKAR, A.; CHUNG, E. Passenger Segmentation Using Smart Card Data. **IEEE Transactions on Intelligent Transportation Systems**, v. 16, n. 3, p. 1537-1548, 2015.

KIM, K. S.; CHEON, S. H.; LIM, S. J. Performance Assessment of Bus Transport Reform in Seoul. **Transportation**, v. 38, p. 719-735, 2011.

KIM, M. K.; KIM, S.; SOHN, H. G. Relationship between Spatio-Temporal travel patterns derived from smart-card data and local environmental characteristics of Seoul, Korea. **Sustain**, v. 10, 2018.

KITAMURA, R. Urban Activity Patterns. *In: INTERNATIONAL Encyclopedia of the Social & Behavioral Sciences*, Amsterdam: ScienceDirect, 2001. p. 15995-15999. Disponível em: <https://www.sciencedirect.com/topics/computer-science/travel-pattern>. Acesso em: 20, jan. de 2021.

KITCHIN, R. Big data and human geography: Opportunities, challenges and risks. **Dialogues in Human Geography**, v. 3, n. 3, p. 262-267, 2013.

KODINARIYA, T. M.; MAKWANA, P. R. Review on determining number of Cluster in *K-means* Clustering. **International Journal of Advance Research in Computer Science and Management Studies**, v. 1, n. 6, p. 90-95, 2013.

KOHONEN, T. **Self-Organizing Map**. 3. ed. ampl. Springer: Berlin, Heidelberg, 2001.

LIMA, L. S. **Espraiamento urbano por autosegregação e seus impactos na acessibilidade urbana de Fortaleza**. 2017. Dissertação (Mestrado em Engenharia de Transportes) – Universidade Federal do Ceará, Fortaleza, 2017.

LITMAN, T. Developing Indicators for Comprehensive and Sustainable Transport Planning. **Transportation Research Record: Journal of the Transportation Research Board**, v. 2017, n. 1, p. 10-15, 2007.

LIU, L.; HOU, A.; BIDERMAN, A.; RATTI, C.; CHEN, J. Understanding individual and collective mobility patterns from smart card records: A case study in Shenzhen. *In: IEEE Conference on Intelligent Transportation Systems, 2009. Proceedings [...]. ITSC, 2009.* p. 842-847.

LONG, J.; NELSON, T.; WULDER, M., Regionalization of Landscape Pattern Indices Using Multivariate *Cluster* Analysis. **Environmental Management**, v. 46, p. 134-142, 2010.

LOPES, A. S. **Transportes, Uso do Solo e Atividades – Modelagem Conceitual para o Planejamento da Acessibilidade Urbana**. 2015. Tese (Doutorado em Engenharia de Transporte) – Universidade Federal do Ceará, Fortaleza, 2015.

LOPES, A. S.; CAVALCANTE, C. B.; LOUREIRO, C. F. G. A Evolução do Planejamento Urbano da Cidade de Fortaleza (Ceará, Brasil) sob o Paradigma do Planejamento Urbano Integrado. *In: PLURIS*, 8., 2018. **Atas [...]**. 2018, p. 12.

LOPES, A. S.; LOUREIRO, C. F. G.; VAN WEE, B. LUTI operational models review based on the proposition of an a priori ALUTI conceptual model. **Transport Reviews**, p. 1-22, 2018.

MA, X.; WU, Y. J.; WANG, Y.; CHEN, F.; LIU, J. Mining smart card data for transit riders' travel patterns. **Transportation Research Part C: Emerging Technologies**, v. 36, n. 2013, p. 1-12, 2013.

MACÁRIO, R. Access as a social good and as an economic good: is there a need of paradigm shift? *In: SCLAR, E. et al. (ed.). Urban Access for the 21st Century, Finance and Governance Models for Transport Infrastructures*. New York, 2014. p. 87-115.

MAGALHÃES, M. T. Q. **Metodologia para Desenvolvimento de Sistemas de Indicadores: Uma Aplicação no Planejamento e Gestão da Política Nacional de Transportes**. 2004. Dissertação (Mestrado) – Departamento de Engenharia Civil e Ambiental, Universidade de Brasília, DF, 2004.

MARTENS, K. Basing Transport Planning on Principles of Social Justice. **Berkeley Plan Journal**, v. 19, p. 1-17, 2006.

MENEZES, E. D. A. G. (2015). **Metodologia para Avaliação Estratégica da Problemática da Acessibilidade Urbana sob o Princípio da Equidade**. 2015. Dissertação (Mestrado em Engenharia de Transportes) – Universidade Federal do Ceará, Fortaleza, 2015.

METROFOR. Metrô de Fortaleza. **Estudo de integração**. 1999.

MEURS, H.; VAN WEE, B. Land use and mobility: a synthesis of findings and policy implications. **Eur. J. Transp. Infrastruct. Res.**, v. 3, n. 2, p. 219-233, 2003.

MEYER, L. H.; ROSER, D. Enough for the future. *In: GOSSERIES, A.; MEYER, L. H. (ed.). Intergenerational Justice*. Oxford: Oxford University Press, 2009.

MEYER, M. D.; MILLER, E. J. **Transportation Planning and Decision Making. Urban transportation planning: a decision-oriented approach**. New York: McGraw-Hill, 2001.

MOKHTARIAN, P. L.; CAO, X. Y. Examining the impacts of residential self-selection on travel behaviour: a focus on methodologies. **Transportation Research B**, v. 42, p. 204-228, 2008.

MORENCY, C.; TREPANIER, M.; AGARD, B. **Analysing the variability of transit users behaviour with smart card data**. *In: IEEE Conference on Intelligent Transportation Systems*, 2009. **Proceedings [...]**. ITSC, 2006. p. 44-49.

MUGION, R. G.; TONI, M.; RAHARJO, H.; DI PIETRO, L.; SEBATHU, S. P. (2018). Does

the service quality of urban public transport enhance sustainable mobility? **Journal of Cleaner Production**, v. 174, p. 1566-1587, 2018.

MUNIZAGA, M. A.; PALMA, C. (2012). Estimation of a disaggregate multimodal public transport Origin-Destination matrix from passive smartcard data from Santiago, Chile. **Transportation Research Part C: Emerging Technologies**, v. 24, p. 9-18, 2012.

NAMRATHA, M.; PRAJWALA, T. R. A comprehensive overview of clustering algorithms in pattern recognition. **IOSR Journal of Computer Engineering (IOSRJCE)**, v. 4, n. 6, p. 23-30, 2012.

NTU. ASSOCIAÇÃO NACIONAL DAS EMPRESAS DE TRANSPORTES URBANOS. **Anuário NTU 2017-2018**, p. 76, 2018.

NUNES, A. A.; DIAS, T. G.; CUNHA, J. F. E. Passenger Journey Destination Estimation from Automated Fare Collection System Data Using Spatial Validation. **IEEE Trans. Intell. Transp. Syst.**, v. 17, p. 133-142, 2015.

OLIVEIRA, J. **Análise da evolução da demanda no transporte coletivo por ônibus em Fortaleza utilizando Big Data**. Projeto de Graduação (Graduação em Engenharia de Civil) – Universidade Federal do Ceará, Fortaleza, 2019.

ORTÚZAR, J. D.; WILLUMSEN, L. G. **Modelling Transport**. 4. ed. Chichester: John Wiley and Sons, 2011.

PAU, S. A. **Using Smart Card Technologies to Measure Public Transport Performance: Data Capture and Analysis**. Barcelona: Universitat Politècnica de Catalunya, 2014.

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

PIERONI, C. **Analysis of travel patterns from precarious settlements transit users in São Paulo through smart card data mining**. 2018. Dissertação (Mestrado em Engenharia de Transportes) – Universidade de São Paulo, São Paulo, 2018.

PINTO, D. G. **Compreensão das desigualdades socioespaciais na acessibilidade ao trabalho formal e informal em Fortaleza**. 2020. Dissertação (Mestrado em Engenharia de Transportes) – Universidade Federal do Ceará, Fortaleza, 2020.

PINTO-CORREIA, T.; BREMAN, B.; JORGE, V.; DNEBOSKÁ, M. **Estudo sobre o Abandono em Portugal Continental**: Análise das dinâmicas da Ocupação do Solo, do Sector Agrícola e da Comunidade Rural – Tipologias de áreas rurais. Évora: Departamento de Planeamento Biofísico e Paisagístico, Universidade de Évora, 2006.

ROKACH, L.; MAIMON, O. Clustering Methods. **Data Mining and Knowledge Discovery Handbook**, p. 321-352, 2015.

- SALOMON, I.; MOKHTARIAN, P. L. What happens when mobility-inclined market segments face accessibility-enhancing policies?. **Transportation Research Part D: Transport Environment**, v. 3, n. 3, p. 129-140, 1998.
- SEABORN, C.; ATTANUCCI, J.; WILSON, N. Analyzing Multimodal Public Transport Journeys in London with Smart Card Fare Payment Data. **Transportation Research Record: Journal of the Transportation Research Board**, n. 2121, p. 119-126, 2009.
- SKUPIN A.; HAGELMAN, R. Visualizing demographic trajectories with self-organizing maps. **GeoInformatica**, v. 9, p. 159-179, 2005.
- SMART, M.; MILLER, M. A.; TAYLOR, B. D. Transit Stops and Stations: Transit Managers' Perspectives on Evaluating Performance. **J. Public Transp.**, 12, p. 59-78, 2009.
- SOARES, F. D. P. **Metodologia de Diagnóstico das Relações entre Uso do Solo e Transportes na Problemática do Planejamento Urbano Integrado**. 2014. Dissertação (Mestrado em Engenharia de Transportes) –Universidade Federal do Ceará, Fortaleza, 2014.
- SONG, Y. C.; MENG, H. D.; GRADY, M. O.; HARE, G. O. HARE. (2007). Applications of Attributes Weighting in Data Mining. *In: IEEE – CONFERENCE ON CYBERNETIC SYSTEMS*, 6., 2007, Dublin. **Proceedings** [...]. Dublin, 2007. p. 41-45.
- SOUSA, F. **Diagnostico estratégico das desigualdades socioespaciais na acessibilidade ao trabalho em Fortaleza**. 2019. Dissertação (Doutorado em Engenharia de Transportes) – Universidade Federal do Ceará, Fortaleza, 2019.
- SOUSA, F. F. L. de M.; MESQUITA, K. G. de A.; LOUREIRO, C. F. G. Caracterização da Evolução do Padrão de Mobilidade de Fortaleza a partir da Calibração do TRANUS. *In: CONGRESSO DA ANPET*, 33., 2019. **Anais** [...]. 2019. p. 2442-2449.
- TERTOOLEN, G.; VAN KREVELD, D.; VERSTRATEN, B. Psychological resistance against attempts to reduce private car use. **Transportation Research Part A: Policy and Practice**, v. 32, n. 3, p. 171-181, 1998.
- TRÉPANIER, M.; MORENCY, C.; AGARD, B. Calculation of Transit Performance Measures Using Smartcard Data. **Journal of Public Transportation**, v. 12, n. 1, p. 79-96, 2009.
- UTENG, T. P. Social sustainability in the transport sector: an essential requisite for sustainable mobility. **International Journal of Environment and Sustainable Development**, v. 6, n. 2, p. 113, 2007.
- UTSUNOMIYA, M.; ATTANUCCI, J.; WILSON, N. H. Potential Uses of Transit *Smart Card* Registration and Transaction Data to Improve Transit Planning. **Transp. Res. Rec. J. Transp. Res. Board**, v. 1971, p. 119-126, 2006.

VALLANCE, S.; PERKINS, H. C.; DIXON J. E. What is social sustainability? A clarification of concepts. **Geoforum**, v. 42, p. 342-348, 2011.

VAN ACKER, V.; VAN WEE, B.; WITLOX, F. When Transport Geography Meets Social Psychology: Toward a Conceptual Model of Travel Behaviour. **Transport Reviews**, v. 30, n. 2, p. 219-240, 2010.

VANDEGINSTE, B. G. M.; MASSART, D. L.; BUYDENS, L. M. C.; DE JONG, S.; LEWI, P. J.; SMEYERS-VERBEKE, J. Cluster analysis. **Handbook of Chemometrics and Qualimetrics: Part B**, p. 57-86, 1998.

VAN WEE, B. Land use and transport: research and policy challenges. **Journal of Transport Geography**, v. 10, p. 259-271, 2002.

VELICKOV, S.; SOLOMATINE, D. Predictive data mining: practical examples. *In: JOINT WORKSHOP ON APPLIED AI IN CIVIL ENGINEERING*, 2., 2000, Cottbus. **Proceedings [...]**. Cottbus, 2000.

WANG, Y.; WANG, Y.; CHOUDHURY, C. Modelling heterogeneity in behavioral response to peak-avoidance policy utilizing naturalistic data of Beijing subway travelers. **Transportation Research Part F: Traffic Psychology and Behaviour**, v. 73, p. 92-106, 2020.

WEISS, E. B. (1992). In *Fairness To Future Generations and Sustainable Development*. **American University International Law Review**, v. 8, n. 1, p. 19-26, 1992.

WITTEN I. E.; FRANK. E. **Data Mining: Practical Machine Learning Tools and Techniques**. 2. ed. Burlington: Morgan Kaufmann, 2005.

XU, Y.; FU, X.; LI, H.; DONG, G.; WANG, Q. A *K-means* Algorithm Based on Feature Weighting. **MATEC Web of Conferences**, v. 232, p. 03005, 2018.

ZAKI, M. J.; MEIRA JR, W. **Data mining and analysis: fundamental concepts and algorithms**. Cambridge: Cambridge University Press, 2014.

ZHAO, J. *et al.* Spatio-Temporal Analysis of Passenger Travel Patterns in Massive Smart Card Data. **IEEE Transactions on Intelligent Transportation Systems**, v. 18, n. 11, p. 3135-3146, 2017.

ZHOU, J.; MURPHY, E.; LONG, Y. Commuting efficiency in the Beijing metropolitan area: an exploration combining smartcard and travel survey data. **Journal of Transport Geography**, v. 41, p. 175-183, 2014.

ZHAO, J.; QU, Q.; ZHANG, F.; XU, C.; LIU, S. Spatio-Temporal Analysis of Passenger Travel Patterns in Massive Smart Card Data. **IEEE Transactions on Intelligent Transportation Systems**, v. 18, n. 11, p. 3135-3146, 2017.

APÊNDICE

Da figura 41 a 49, são os indicadores para o ano de 2014. Da figura 50 a 52, são os indicadores para o ano de 2018. Da figura 53 a 55, são os indicadores para evolução entre os anos 2014 e 2018.

Figura 41 – Mapa de quintis; Lisa Map e Boxplot para o indicador de frequência média por padrão para o ano de 2014

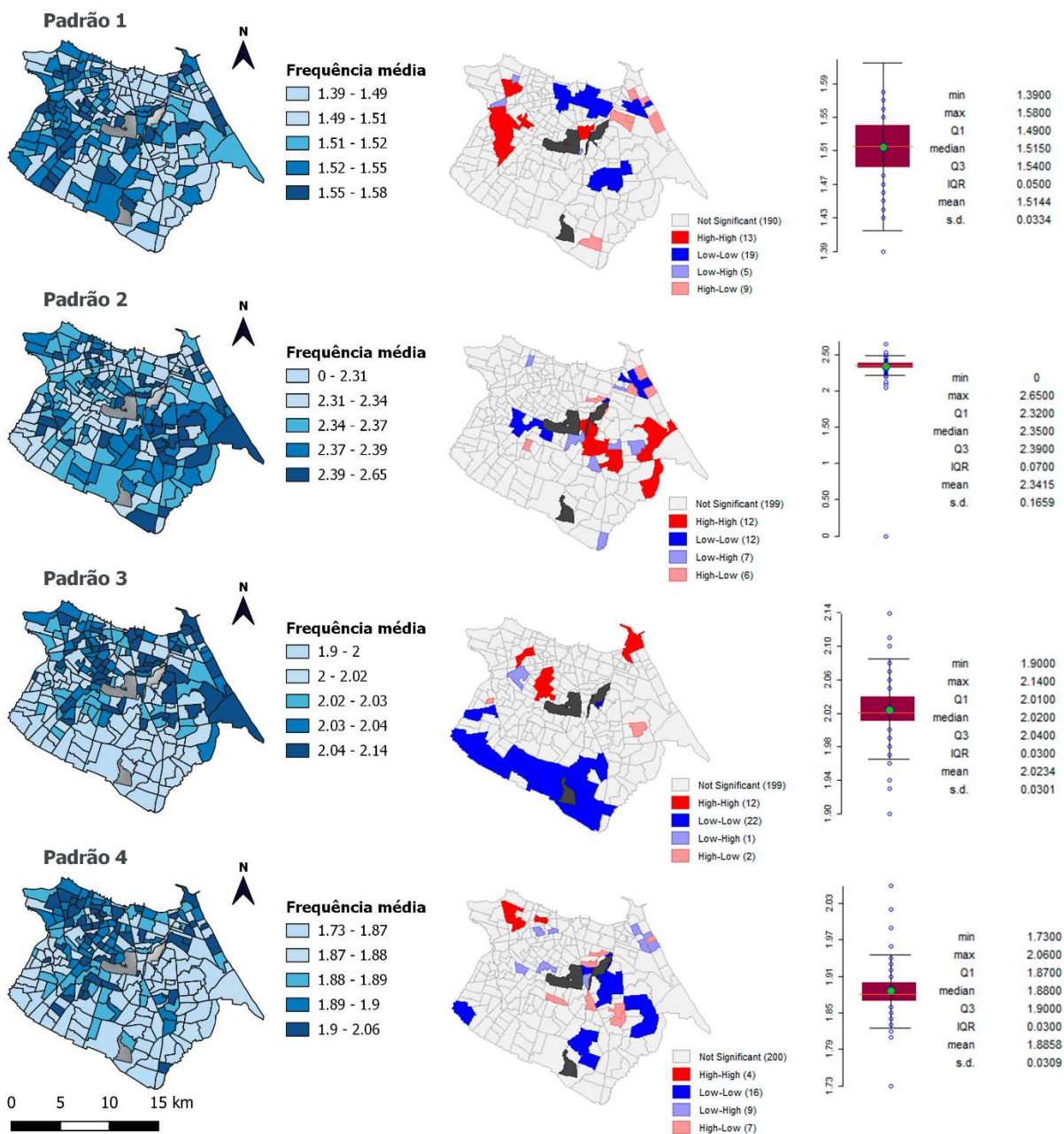
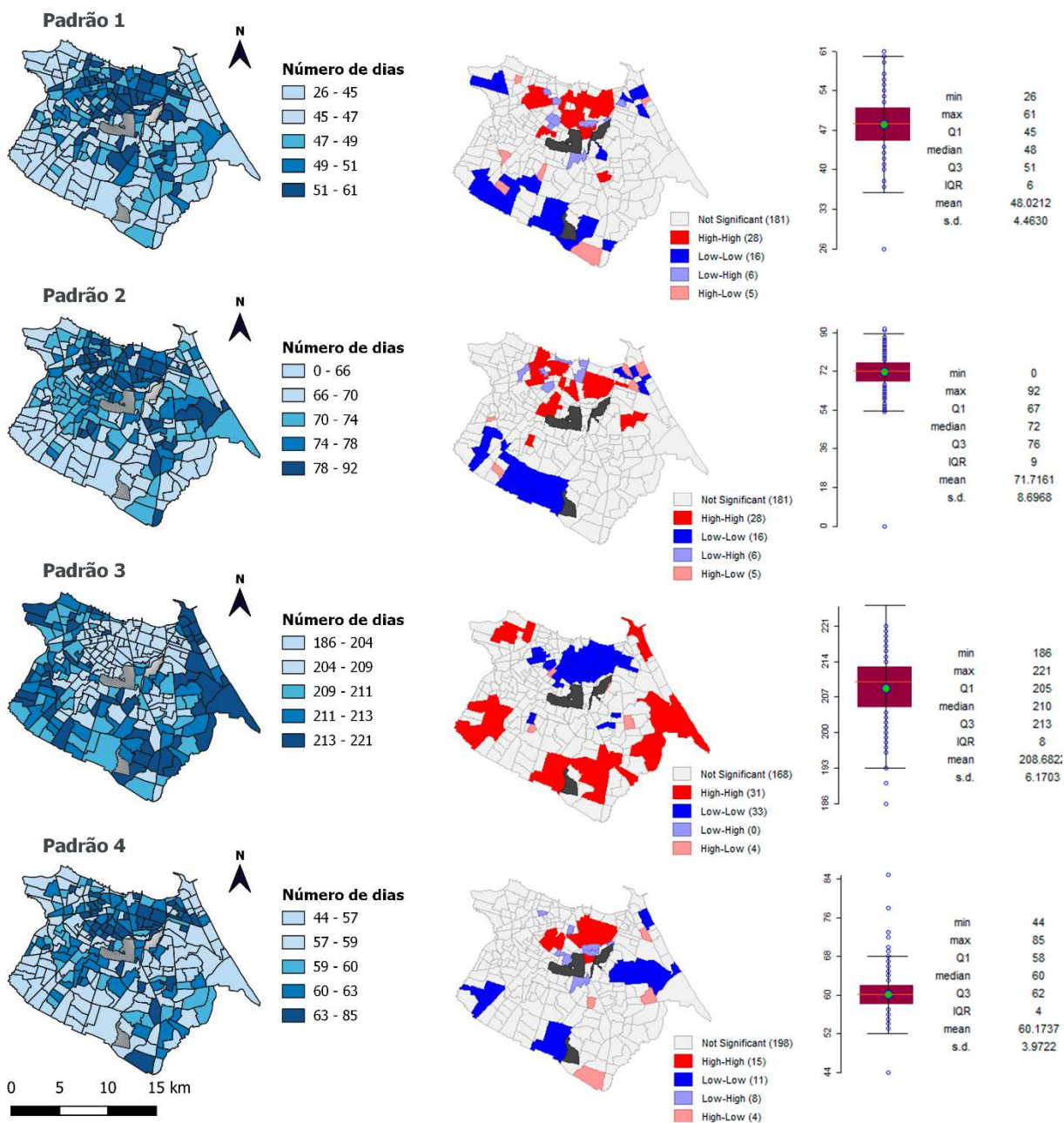
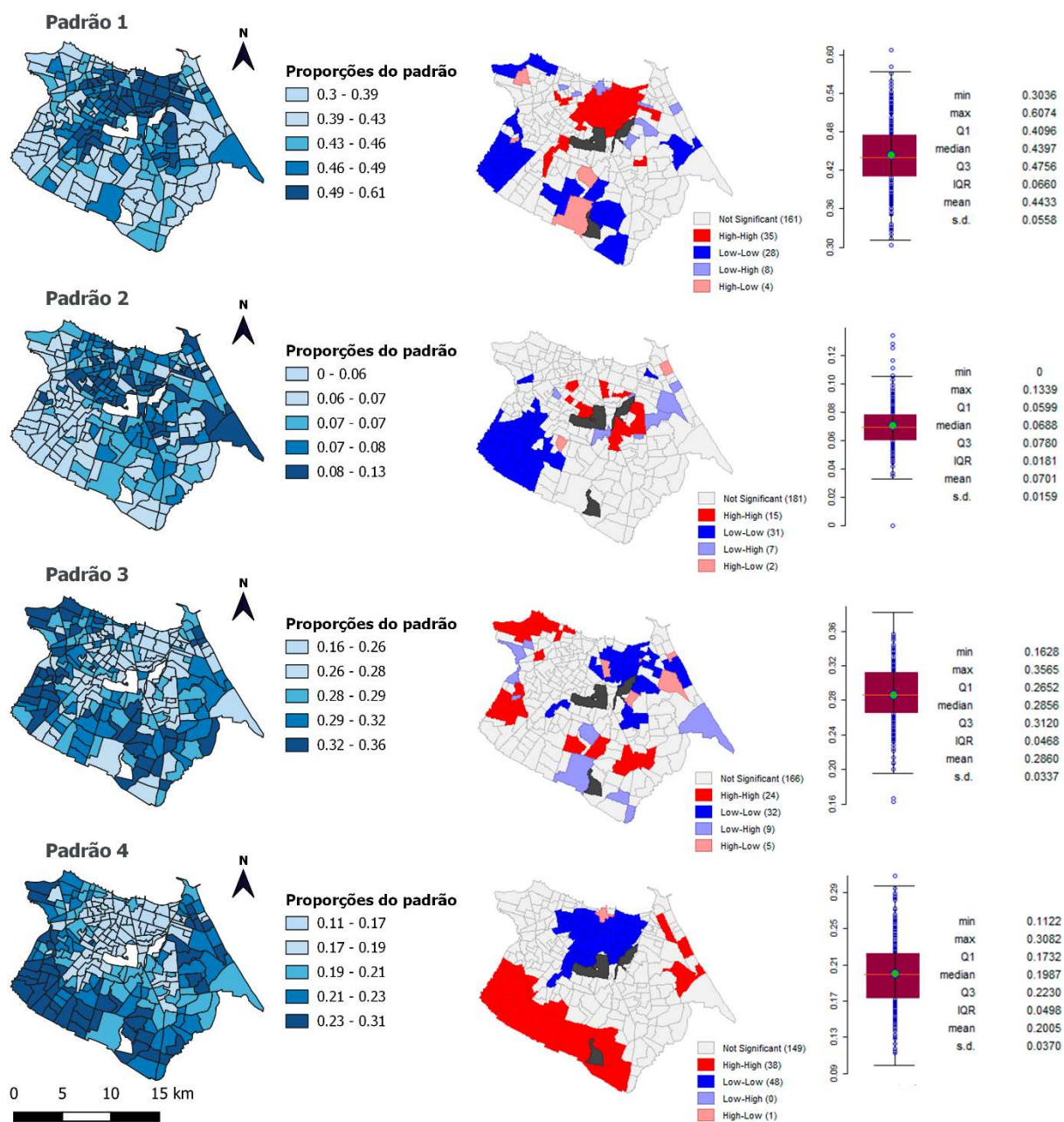


Figura 42 – Mapa de quintis; Lisa Map e Boxplot para o indicador de número de dias de uso por padrão para o ano de 2014



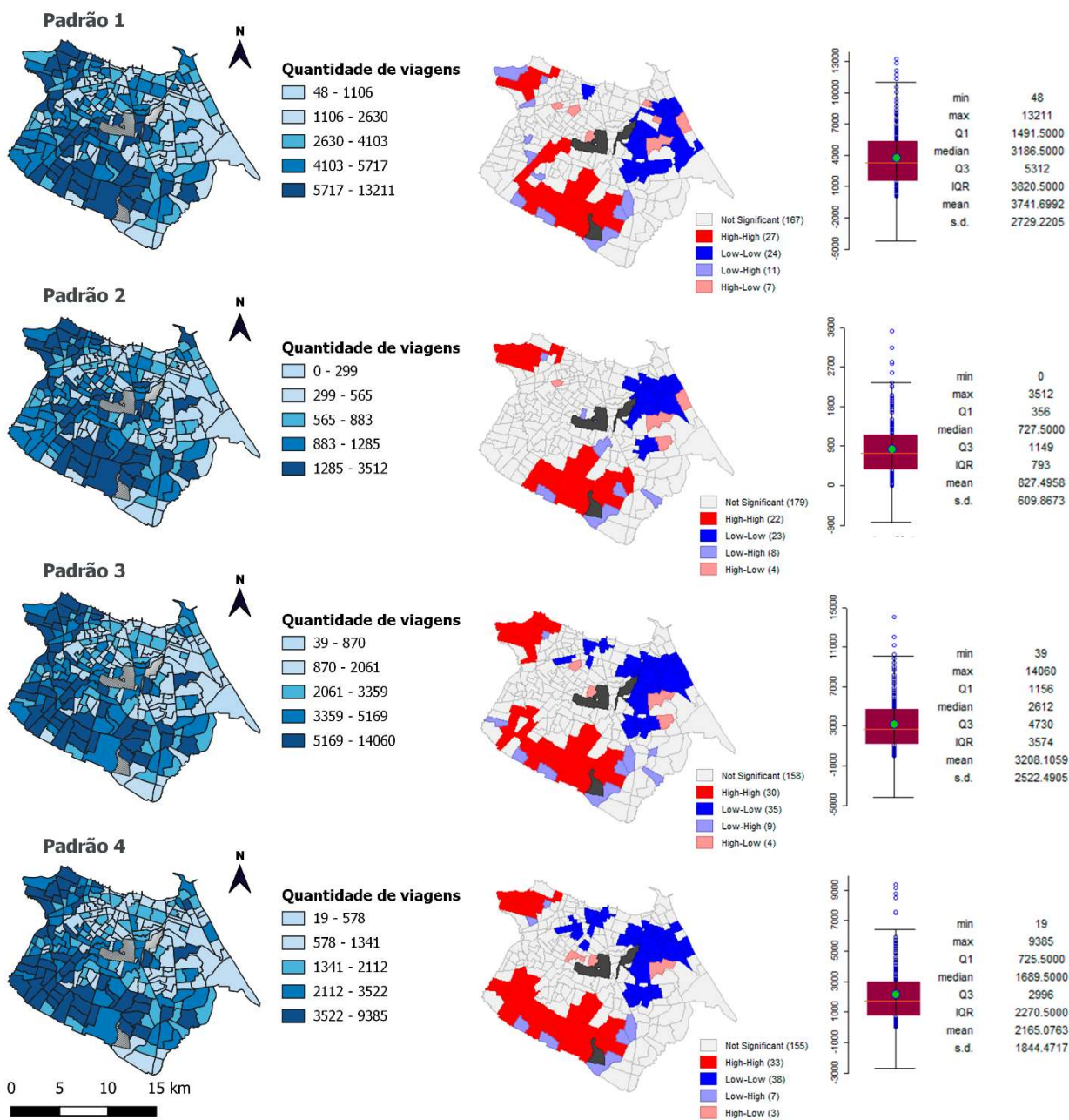
Fonte: Elaborado pelo autor.

Figura 43 – Mapa de quintis; Lisa Map e Boxplot para o indicador de proporções dos padrões para o ano de 2014



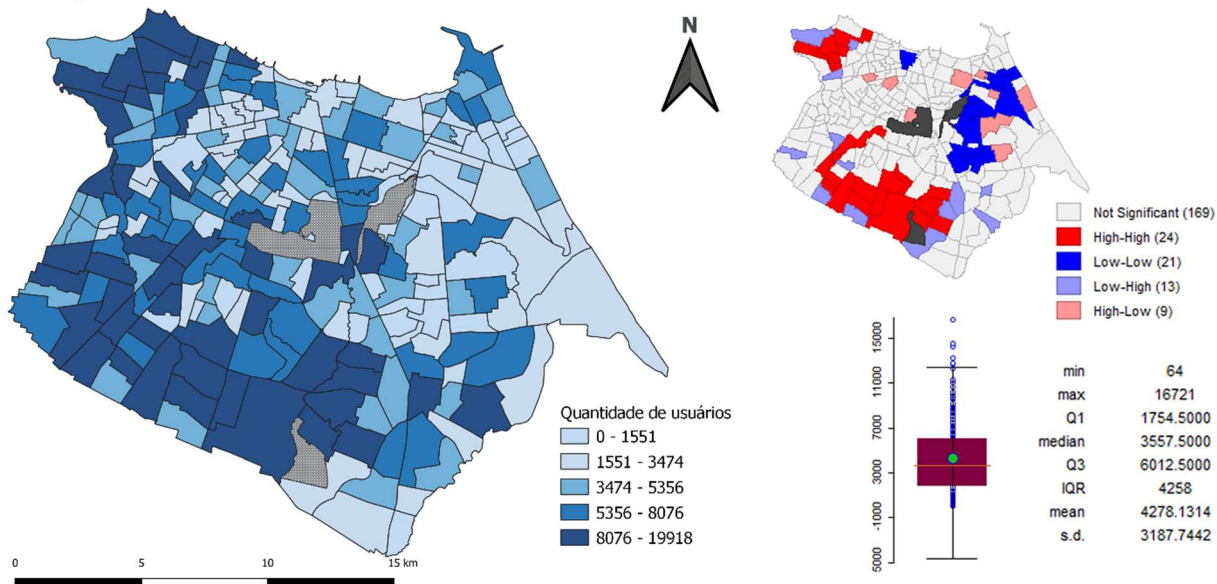
Fonte: Elaborado pelo autor.

Figura 44 – Mapa de quintis; Lisa Map e Boxplot para o indicador viagens produzidas em um dia útil por padrão para o ano de 2014



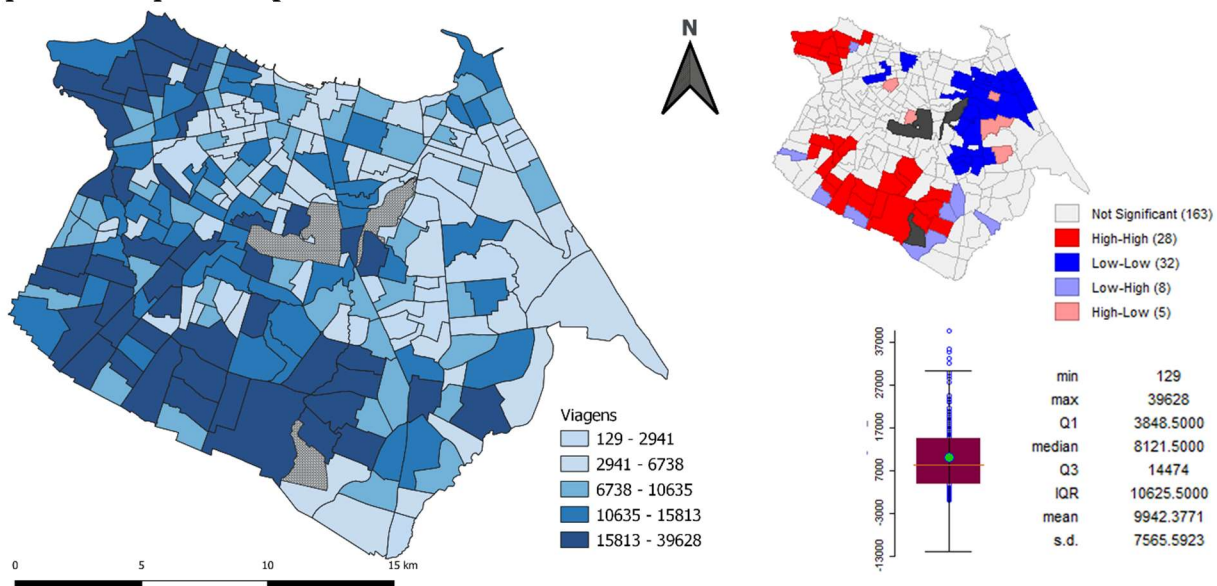
Fonte: Elaborado pelo autor.

Figura 45 – Mapa de quintis; Lisa Map e Boxplot para o indicador quantidade de usuários por zona para o ano de 2014



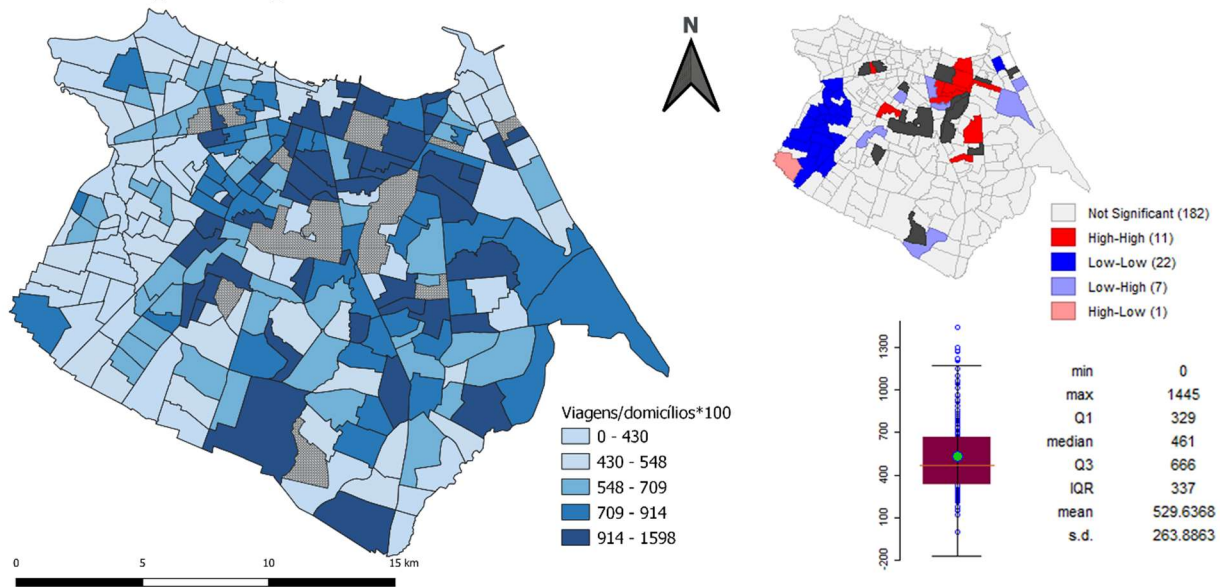
Fonte: Elaborado pelo autor.

Figura 46 – Mapa de quintis; Lisa Map e Boxplot para o indicador quantidade de viagens produzidas por zona para o ano de 2014



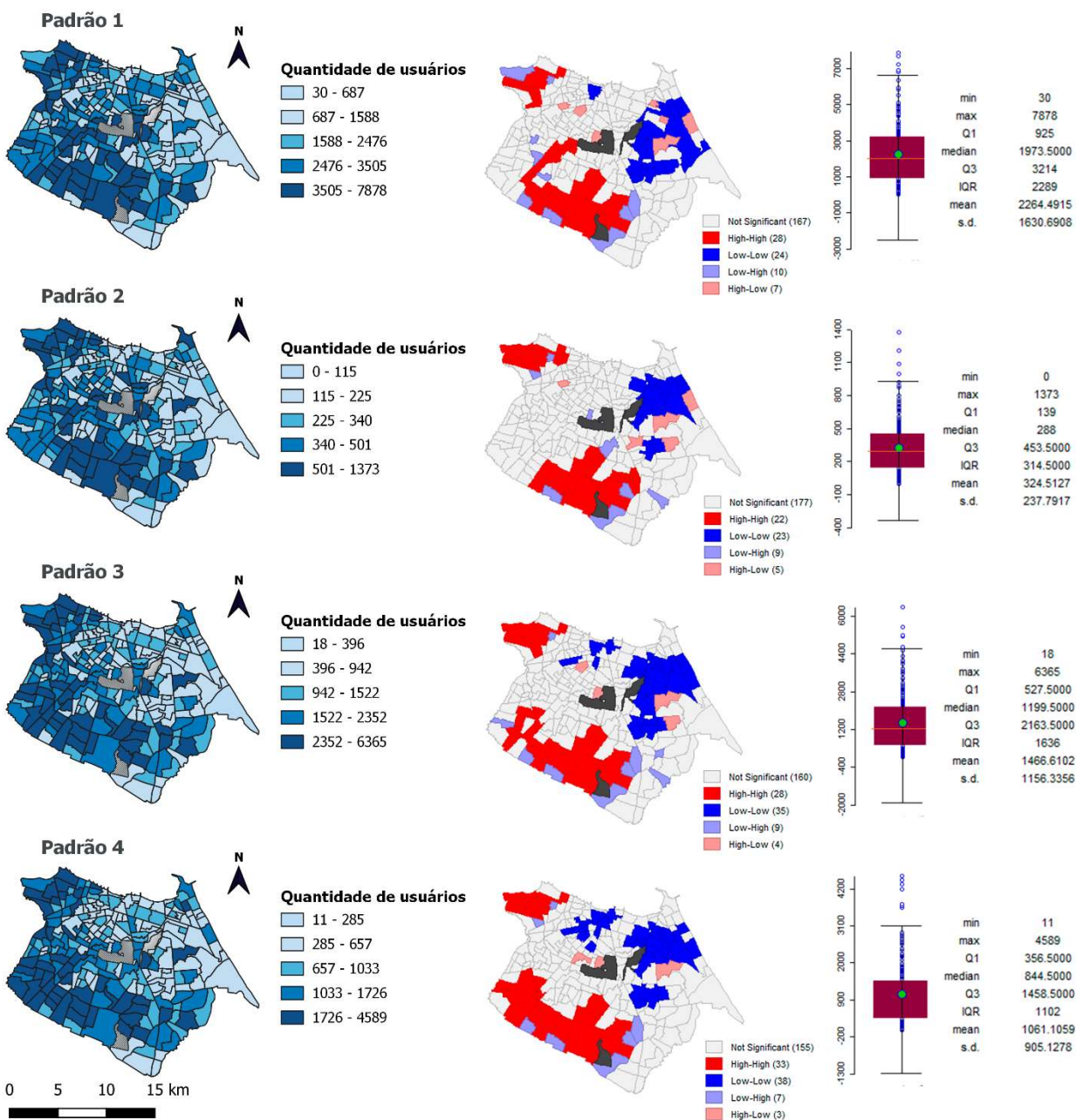
Fonte: Elaborado pelo autor.

Figura 47 – Mapa de quintis; Lisa Map e Boxplot para o indicador quantidade de viagens por domicílios por zona para o ano de 2014



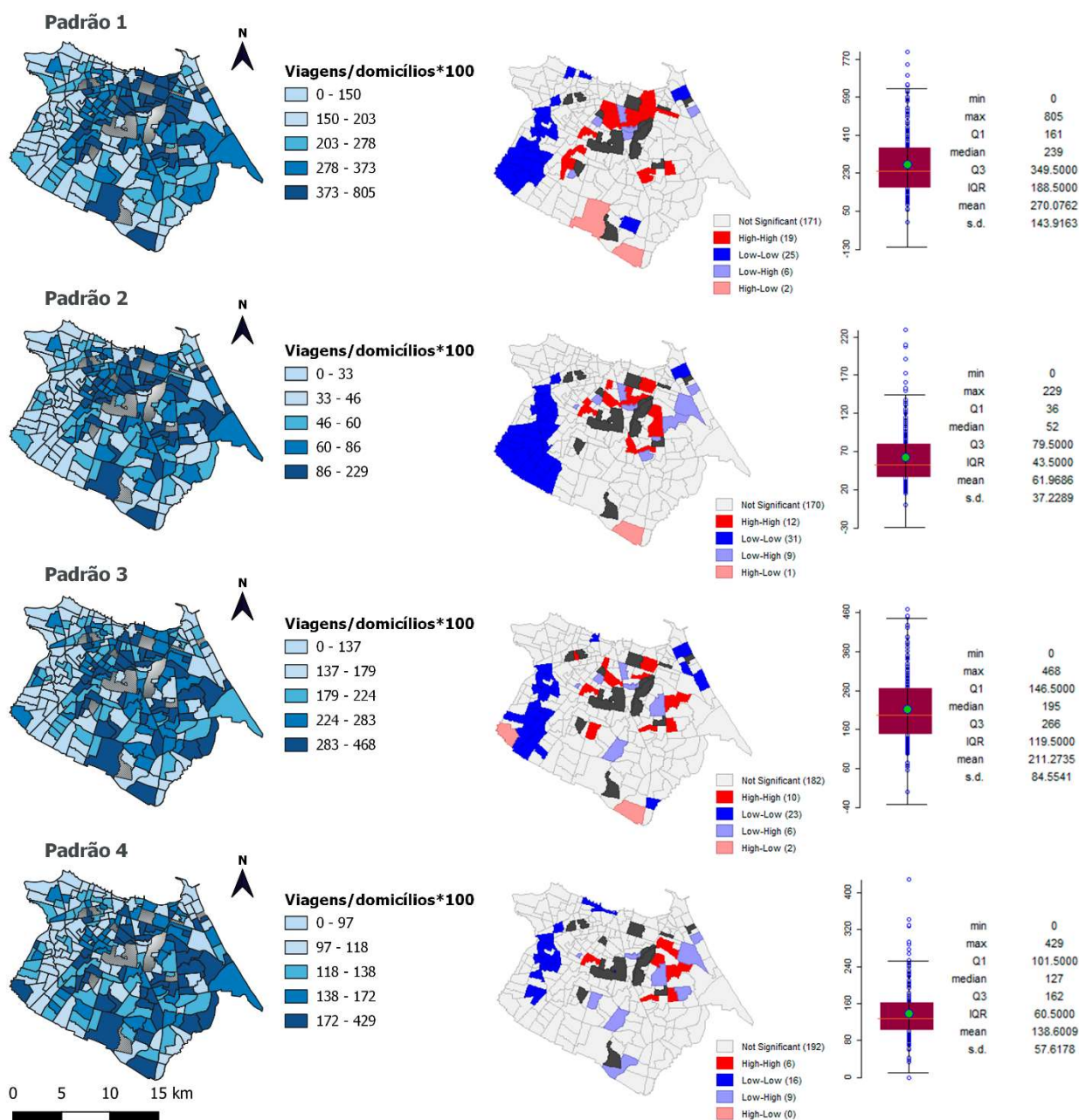
Fonte: Elaborado pelo autor.

Figura 48 – Mapa de quintis; Lisa Map e Boxplot para o indicador quantidade de usuários por padrão para o ano de 2014



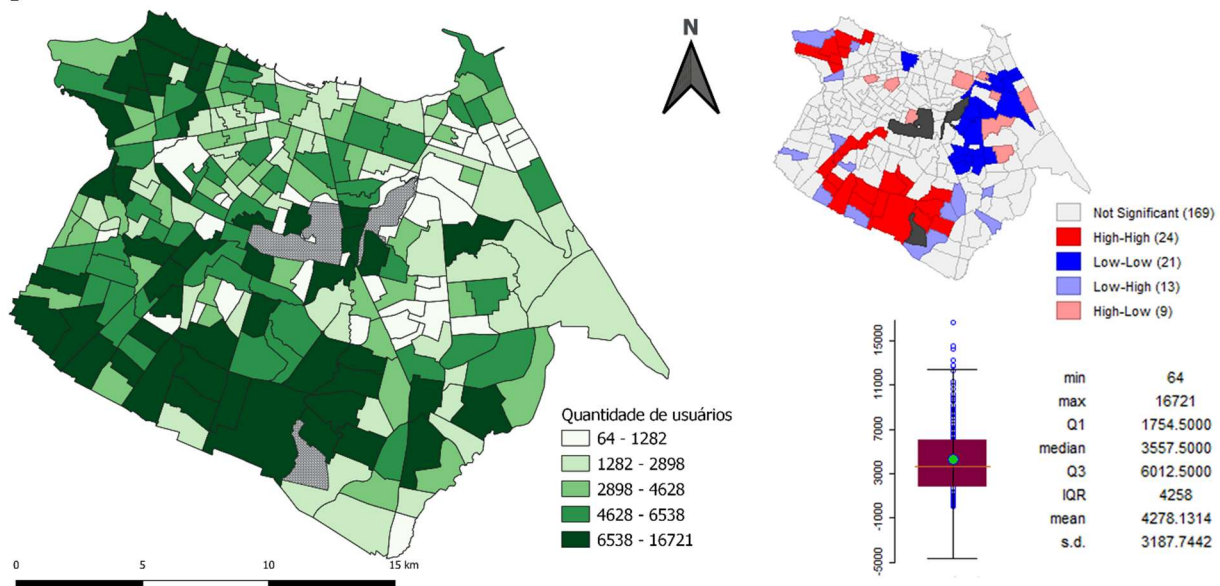
Fonte: elaborado pelo autor.

Figura 49 – Mapa de quintis; Lisa Map e Boxplot para o indicador viagens por domicílios por padrão para o ano de 2014



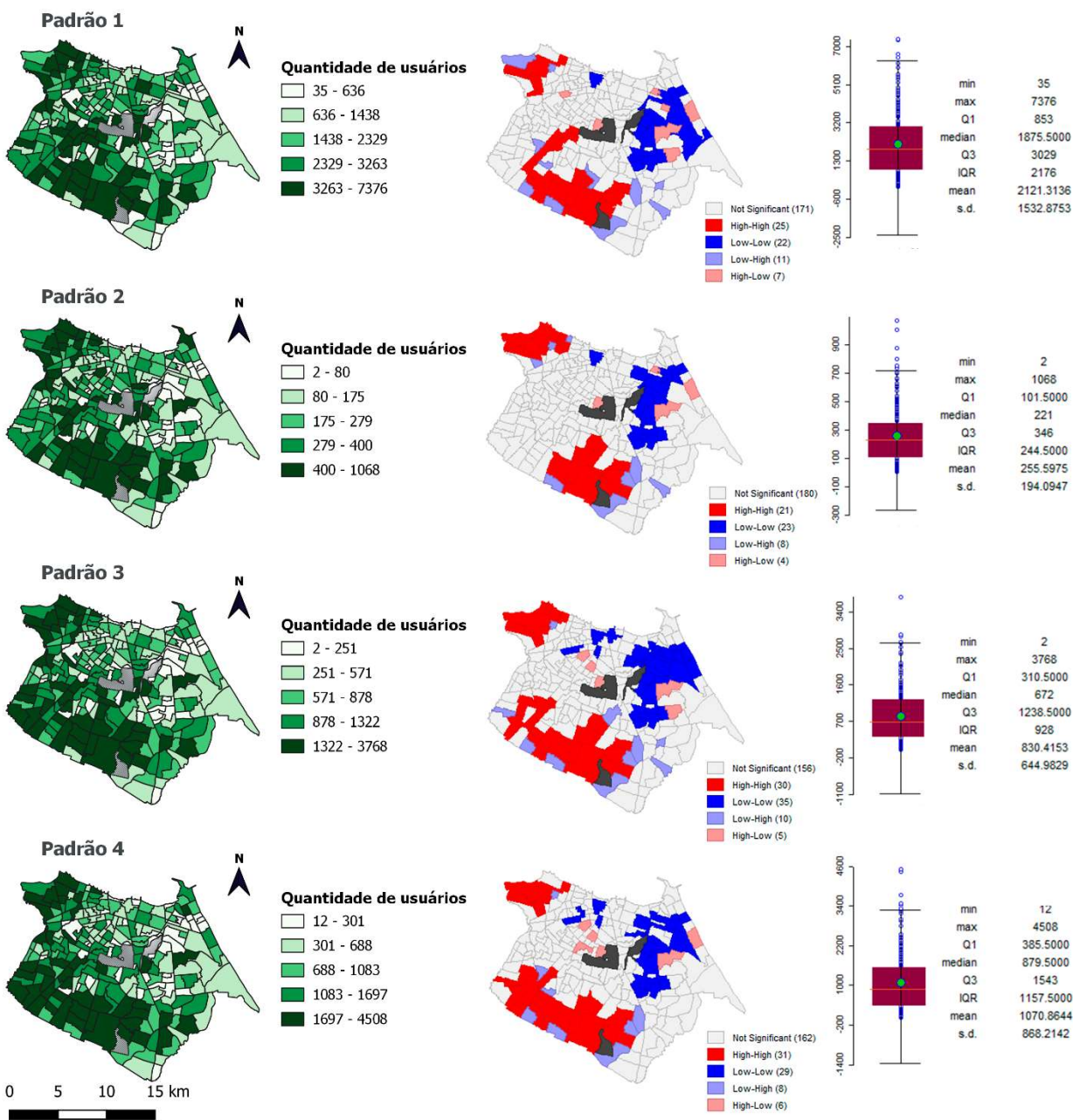
Fonte: elaborado pelo autor.

Figura 50 – Mapa de quintis; Lisa Map e Boxplot para o indicador quantidade de usuários zona para o ano de 2014



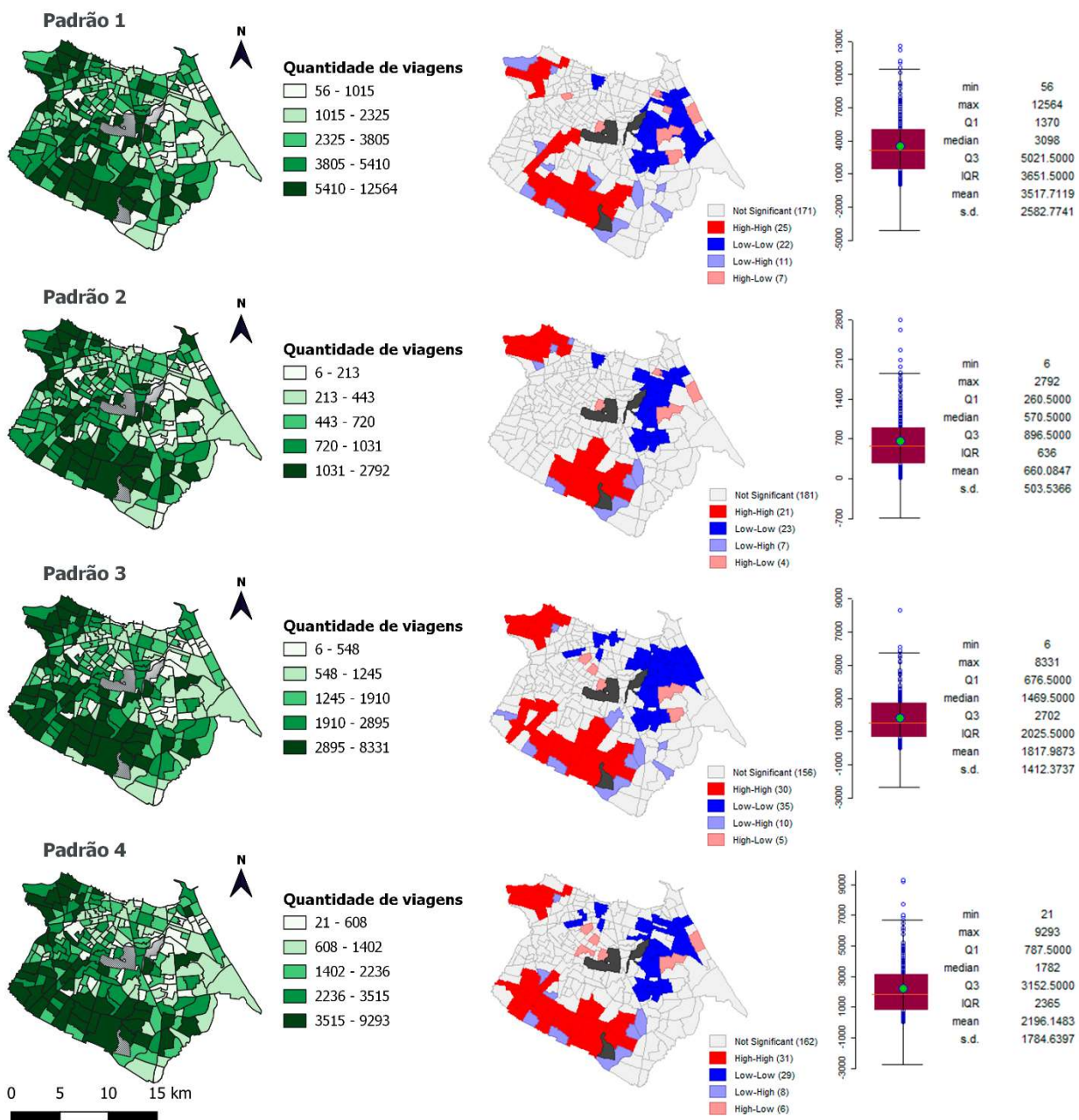
Fonte: Elaborado pelo autor.

Figura 51 – Mapa de quintis; Lisa Map e Boxplot para o indicador quantidade de usuários por padrão para o ano de 2018



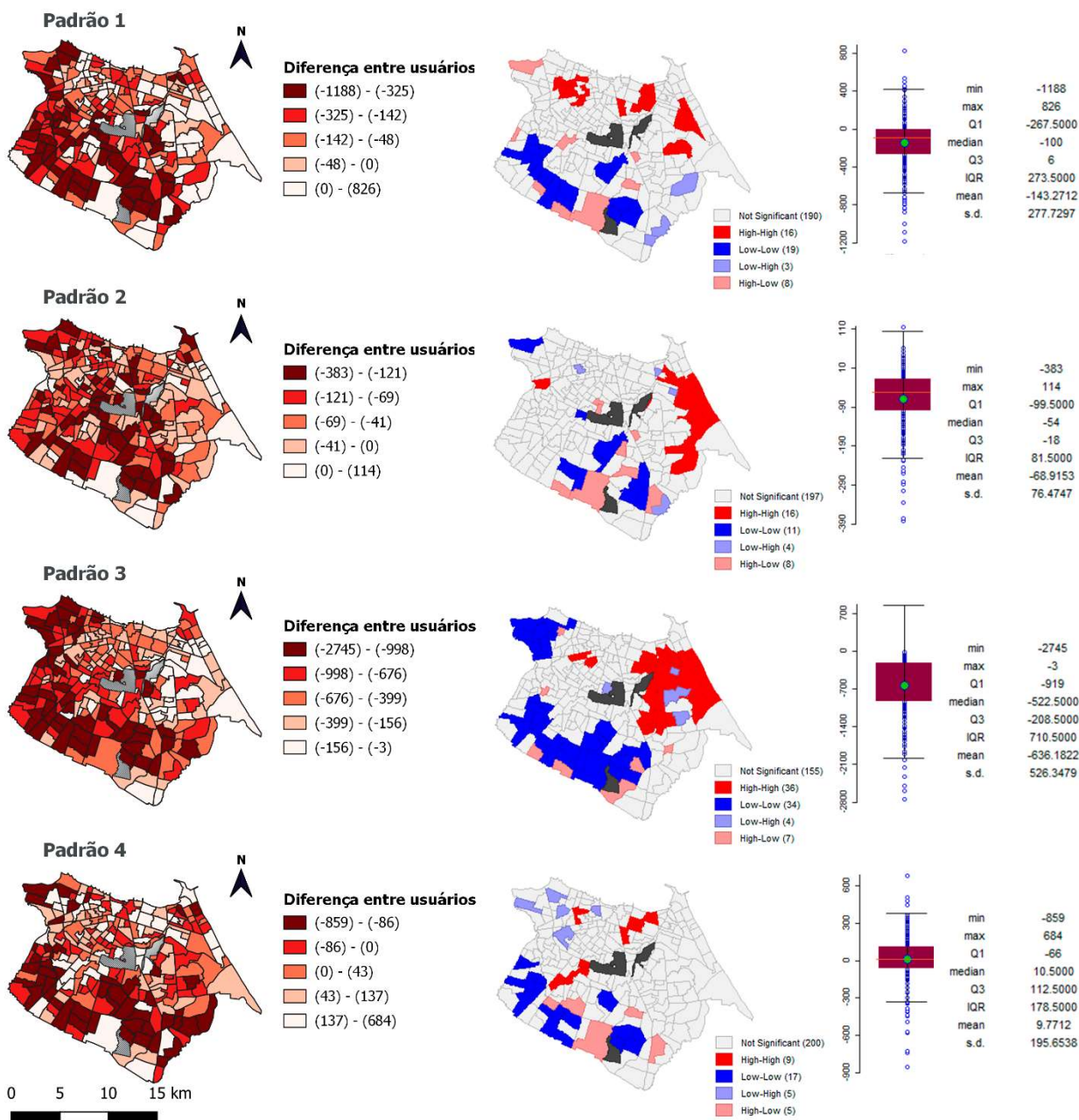
Fonte: elaborado pelo autor.

Figura 52 – Mapa de quintis; Lisa Map e Boxplot para o indicador viagens produzidas em um dia útil por padrão para o ano de 2018



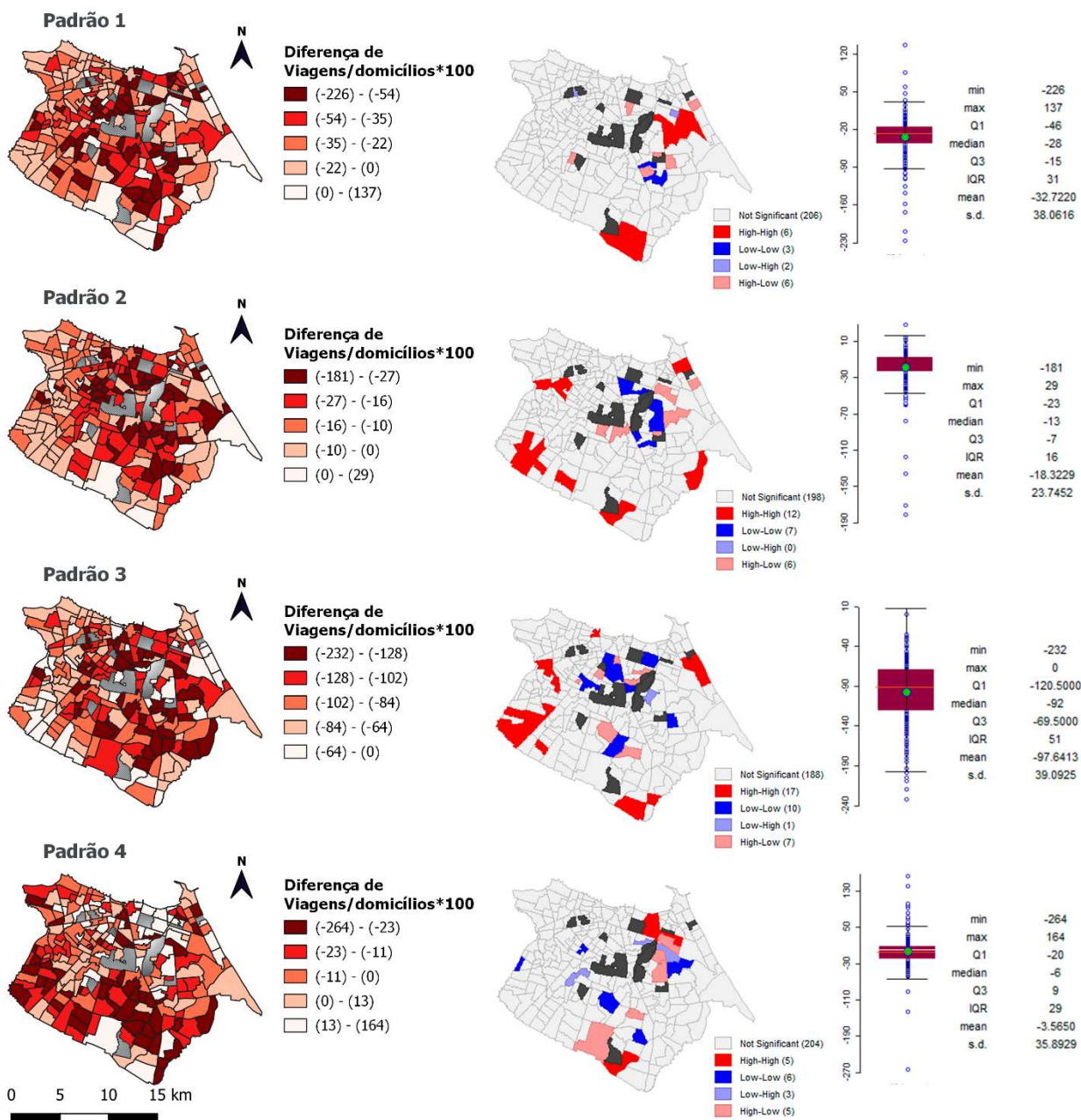
Fonte: elaborado pelo autor.

Figura 53 – Mapa de quintis; Lisa Map e Boxplot para o indicador diferença de usuários por padrão



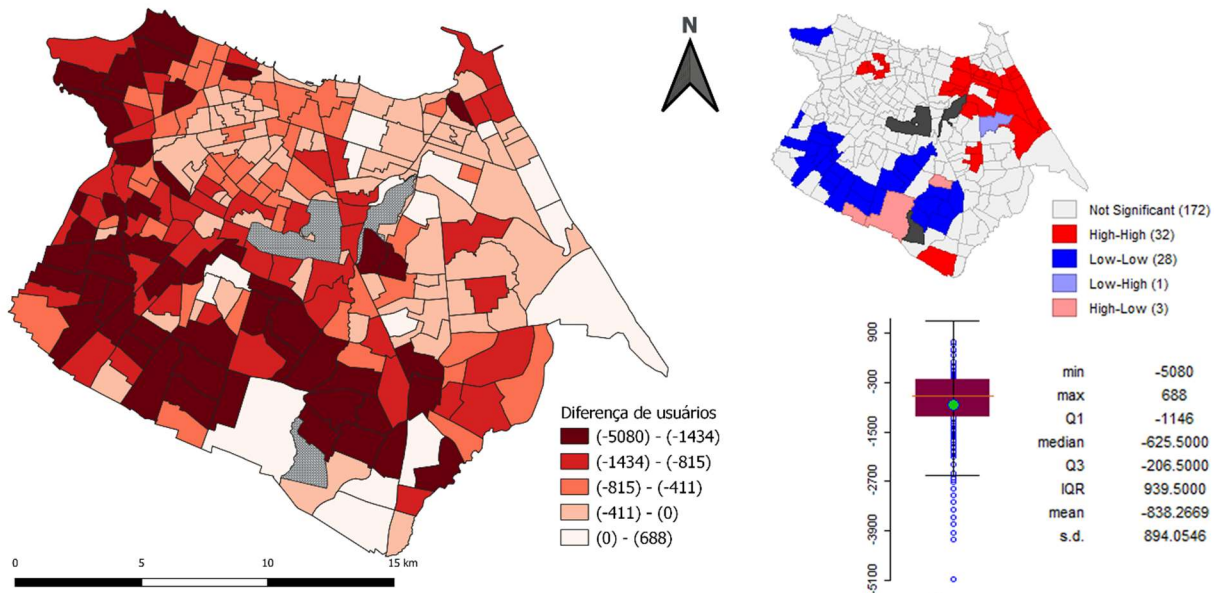
Fonte: elaborado pelo autor.

Figura 54 – Mapa de quintis; Lisa Map e Boxplot para o indicador diferença de viagens por domicílios por padrão



Fonte: elaborado pelo autor.

Figura 55 – Mapa de quintis; Lisa Map e Boxplot para o indicador diferença de usuários por domicílios



Fonte: elaborado pelo autor.