UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
PROGRAMA DE PÓS -GRADUAÇÃO EM COMPUTAÇÃO APLICADA

YUSSEF PARCIANELLO

ANÁLISE DE ORIGEM-DESTINO DO USO DO SISTEMA DE TRANSPORTE COLETIVO DE CURITIBA SOB O PONTO DE VISTA DE *REGIONS OF INTEREST*

DISSERTAÇÃO DE MESTRADO

CURITIBA
2020

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
PROGRAMA DE PÓS -GRADUAÇÃO EM COMPUTAÇÃO APLICADA

YUSSEF PARCIANELLO

ANÁLISE DE ORIGEM-DESTINO DO USO DO SISTEMA DE TRANSPORTE COLETIVO DE CURITIBA SOB O PONTO DE VISTA DE *REGIONS OF INTEREST*

Dissertação submetida ao Programa de Pós-Graduação em Computação Aplicada da Universidade Tecnológica Federal do Paraná como requisito para a obtenção do título de Mestre em Computação Aplicada.

Área de concentração: Engenharia de Sistemas Computacionais

Orientador: Nádia Puchalski Kozievitch
Co-orientador: Keiko Veronica Ono Fonseca

CURITIBA
2020

Dados Internacionais de Catalogação na Publicação

Parcianello, Yussef

Análise de origem-destino do uso do sistema de transporte coletivo de Curitiba sob o ponto de vista de *Regions of Interest* [recurso eletrônico] / Yussef Parcianello. -- 2019.

1 arquivo eletrônico (88 f.) : PDF ; 5,10 MB.

Modo de acesso: World Wide Web.

Texto em português com resumo em inglês.

Dissertação (Mestrado) - Universidade Tecnológica Federal do Paraná. Programa de Pós-graduação em Computação Aplicada. Área de Concentração: Engenharia de Sistemas Computacionais, Curitiba, 2019.

Bibliografia: f. 71-75.

1. Computação - Dissertações. 2. Transporte público urbano - Curitiba (PR). 3. Levantamentos de origem e destino do trânsito - Curitiba (PR) - Avaliação. 4. Visualização da informação. 5. Software - Desenvolvimento. 6. Software gratuito. 7. Interfaces de usuário (Sistemas de computação) - Testes. 8. Sistemas de informação gerencial. 9. Simulação (Computadores).

I. Kozievitch, Nádia Puchalski, orient. II. Fonseca, Keiko Verônica Ono, coorient. III. Universidade Tecnológica Federal do Paraná. Programa de Pós-graduação em Computação Aplicada. IV. Título.

CDD: Ed. 23 -- 621.39

Biblioteca Central do Câmpus Curitiba - UTFPR
Bibliotecária: Luiza Aquemi Matsumoto CRB-9/794

Ministério da Educação
Universidade Tecnológica Federal do Paraná
Diretoria de Pesquisa e Pós-Graduação

TERMO DE APROVAÇÃO DE DISSERTAÇÃO

A Dissertação de Mestrado intitulada **"Análise de origem- destino do uso do sistema de transporte público de Curitiba sob o ponto de vista de Regions of interest"** defendida em sessão pública pelo(a) candidato(a) **Yussef Parcianello**, no dia **19 de dezembro de 2019**, foi julgada para a obtenção do título de Mestre em Ciências, Área de Concentração: **Engenharia de Sistemas Computacionais**, Linha de Pesquisa: **Sistemas de Informação,** e aprovada em sua forma final, pelo Programa de Pós-Graduação em Computação Aplicada.

BANCA EXAMINADORA:

Profa. Dra. Nádia Puchalski Kozievitch - Presidente – UTFPR
Profa. Dra. Maristela Terto de Holanda – UNB
Prof. Dr. Marcelo de Oliveira Rosa – UTFPR
Profa. Dra. Juliana de Santi - UTFPR

A via original deste documento encontra-se arquivada na Secretaria do Programa, contendo a assinatura da Coordenação após a entrega da versão corrigida do trabalho.

Curitiba, 19 de dezembro de 2019.

_____
Carimbo e Assinatura do(a) Coordenador(a) do Programa

Dedico este trabalho à todos aqueles que superaram os desafios que lhes foram impostos de maneira a se tornarem pessoas melhores.

AGRADECIMENTOS

À minha orientadora, professora Nádia Puchalski Kozievitch, pelo apoio, pela disponibilidade, pelas críticas construtivas e pelos direcionamentos.

À minha co-orientadora, professora Keio Veronica Ono Fonseca, também pelo apoio e pelas contribuições dadas ao longo deste trabalho.

À coordenação da Pós, pela atenção e pela colaboração em cada etapa deste estudo.

Aos demais profissionais da educação que direta e/ou indiretamente contribuíram para a realização deste trabalho.

Aos meus familiares pelo carinho, palavras de ânimo, incentivo e pela compreensão nos inevitáveis momentos de ausência.

À minha esposa, Juciane, pela ajuda, apoio incondicional, pelo amor, paciência, e por dar sentido a minha vida.

RESUMO

Yussef Parcianello., ANÁLISE DE ORIGEM-DESTINO DO USO DO SISTEMA DE TRANSPORTE COLETIVO DE CURITIBA SOB O PONTO DE VISTA DE *REGIONS OF INTEREST*. 88 f. Dissertação de Mestrado - Programa de Pós-Graduação em Computação Aplicada. Curitiba - PR, 2020.

O Sistema de Transporte Público (STP) e seu gerenciamento de operações requerem o processamento de grandes volumes de dados (rotas de coletivos, dados de usuários, horários de ônibus, etc.). Tais dados podem oportunizar a identificação de facetas do comportamento do usuário, de necessidades de transporte e de tendências de tráfego. Em particular, dados de origem-destino servem para indicar padrões e escolhas de viagens dos cidadãos, fornecendo subsídios que permitem compreender, inclusive, a dinâmica da ocupação do espaço urbano. Diante deste cenário, este trabalho apresenta um novo protótipo de visualização de dados de origem-destino desenvolvido a partir de tecnologias open-source, não demandando nenhum tipo de gasto referente a aquisição e/ou contratação de licença de *software*. A solução desenvolvida é inteiramente *online* e não exige a configuração ou instalação de nenhum tipo de *software* ou *plugin* para seu funcionamento. As consultas podem ser compostas e disparadas pelo usuário, e tais resultados podem ser exibidos através de diferentes representações visuais e agregados sob diferentes níveis de detalhamento, mantendo sempre os contextos espaciais e temporais. A ferramenta permite compreender aspectos relacionados ao perfil do usuário do transporte público, a variação da demanda do STP ao longo de um determinado período, os locais onde ocorrem a maior quantidade de embarques e desembarques, compreender a dinâmica de ocupação do espaço urbano, dentre outros. As funcionalidades oferecidas na solução foram implementadas a partir de uma lista de requisitos levantada em meio a um grupo de pesquisadores da área de mobilidade urbana. Resultados obtidos mediante realização de um teste de usabilidade indicaram que a solução permite que usuários realizem de forma facilitada análises de origem-destino sem precisar utilizar qualquer tipo de linguagem de programação e/ou de manipulação de dados.

Palavras-chave: Origem-destino. Visualização de Dados. Sistema de Transporte Público. Sistema de Informação Geográfica .

ABSTRACT

PARCIANELLO, Yussef., ORIGIN-DESTINATION ANALYSIS OF THE USE OF CURITIBA'S PUBLIC TRANSPORTATION SYSTEM FROM THE POINT OF VIEW OF REGIONS OF INTEREST. 88 f. Dissertação de Mestrado - Programa de Pós-Graduação em Computação Aplicada. Curitiba - PR, 2020.

The Public Transportation System (PTS) and its operation management require the processing of large amount of data (bus routes, user data, bus schedules, etc.). Such data provides number of opportunities to identify various facets of user behavior, transport needs and traffic trends. In particular, origin-destination data serve to indicate citizens' travel patterns and choices, providing insights related to the dynamic of the urban space occupation. Given this scenario, this paper presents a new prototype of origin-destination data visualization developed using only open-source technologies, not requiring any kind of expenses related to the acquisition and /or contracting of *software* license. The solution is entirely online and does not require the configuration or installation of any type of *software* or *plugin* in order to run. Queries can be composed and triggered by the user, and such results can be displayed through different visual representations and aggregated under different levels of detail, always maintaining the spatial and temporal contexts. The tool allows us to understand aspects related to the profile of users of the public transportation, the variation in demand of PTS over a given period, the places where the greatest number of boarding and landings occur, understand the dynamics of urban space occupation, among others. The functionalities offered in the solution were implemented from a list of requirements raised among a group of urban mobility researchers. Results obtained through a usability test indicated that the solution allows lay users to easily perform source-destination analyzes without having to use any programming language and/or data manipulation.

Keywords: Origin-Destination. Data Visualization. Public Transportation System. Geospatial Information System .

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIAÇÕES

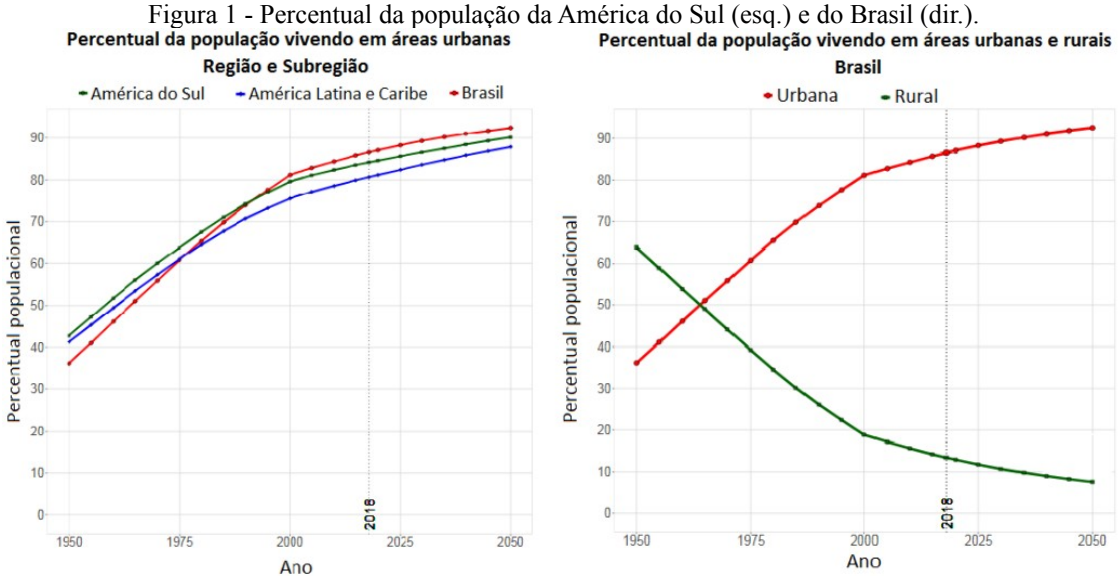
PPGCA	Programa de Pós-Graduação em Computação Aplicada
UTFPR	Universidade Tecnológica Federal do Paraná
OD	Origem-Destino
STP	Sistema de Transporte Coletivo
PTS	Public Transportation System
ROI	*Region of Interest*
RDI	Região de Interesse
POI	Point of Interest
PDI	Ponto de Interesse
SIG	Sistema de Informação Geográfica
GIS	Geospatial Information System
SIGT	Sistema de Informação Geográfica para Transportes
GIST	Geospatial Information System for Transportation
SQL	Structured Query Language
PDA	Portal de Dados Abertos
IPPUC	Instituto de Pesquisa e Planejamento Urbano
URBS	Urbanização de Curitiba
PMC	Prefeitura Municipal de Curitiba
SBEP	Sistema de Bilhetagem Eletrônica de Passagens
SGBD	Sistema Gerenciador de Bancos de Dados

SUMÁRIO

1 INTRODUÇÃO

Vivemos em uma sociedade na qual mais da metade da população mundial reside em cidades. Segundo o estudo realizado pela Organização das Nações Unidas (ONU)[1], na década de 1950, apenas 30% da população mundial vivia em áreas urbanas, um número que cresceu para 55% em 2018. A Figura 1 mostra que cerca de 80% da população sul-americana vive em regiões urbanas: no caso do Brasil, este número sobe para 87%, equivalente a mais de 180 milhões de pessoas vivendo nas cidades. Este aumento da população urbana traz consigo um série de desafios que impactam direta e indiretamente o bem-estar da população. Dentre os desafios relacionados a mobilidade urbana, podemos citar o aumento da frota de veículos, poluição sonora e atmosférica, engarrafamentos, aumento dos tempos de viagens, dentre outros.

Figura 1 - Percentual da população da América do Sul (esq.) e do Brasil (dir.).



Fonte: Adaptado de Nations (2018).

Neste cenário, Schrier (2014) menciona que o movimento de abertura de dados (*Open Data*, em inglês) tem sido uma estratégia utilizada pela administração pública para permitir que a comunidade se engaje na busca por soluções para demandas das cidades. Isto tem sido feito através da oferta de conjuntos de dados via Portais de Dados Abertos (PDA). Cidades como Paris[2], Nova Iorque[3] e Moscou[4] têm buscado disponibilizar tais dados através

1 https://population.un.org/wup/Publications/Files/WUP2018-PopFacts_2018-1.*PDF* - último acesso em Ago. 25, 2019.
2 opendata.paris.fr - último acesso em Mar. 17, 2019.
3 opendata.cityofnewyork.us - último acesso em Mar. 17, 2019.
4 data.gov.ru - último acesso em Mar. 17, 2019.

de seus respectivos PDA. No Brasil, Porto Alegre[5], São Paulo[6], Natal[7], Recife[8] e Distrito Federal[9] são algumas das cidades que disponibilizam dados que abordam diferentes assuntos em seus PDA. Curitiba também tem acompanhado este movimento, disponibilizando dados referentes a mobilidade urbana via prefeitura[10], Instituto de Pesquisa e Planejamento Urbano (IPPUC)[11] e Urbanização de Curitiba (URBS)[12].

Para Freitas et al. (2001), essa crescente oferta de dados abertos *online* traz consigo não só oportunidades de pesquisa, mas também uma série de desafios. Trabalhos como os de Beluzo (2015), Costa et al. (2017), Simette et al. (2018) e Kozievitch et al. (2018) evidenciam alguns dos desafios enfrentados em estudos que envolvem dados abertos, tais como: inexistência (ou imprecisão) de dados e/ou de componentes espaciais, divergência na forma e no nível de detalhamento dos dados disponibilizados, má qualidade (ou ausência) dos metadados, integração de dados disponibilizados em diferentes formatos de arquivos, integração de dados georreferenciados disponibilizados em diferentes sistemas de referência e a integração de dados disponibilizados por diferentes provedores, contendo dados do mesmo assunto, mas em diferentes níveis de detalhamento e/ou de precisão.

Conforme a Associação Nacional De Transportes Públicos (1997), as relações entre a ocupação do solo e a infraestrutura de transporte explicam e condicionam o desenvolvimento das cidades. Nesta direção, planejar o espaço urbano significa, muitas vezes, ter de focar em programas relacionados ao sistema de transporte e sua infraestrutura, visto ser este um dos principais ordenadores do espaço. Nesta direção, estudos de Origem-Destino (OD) representam um importante instrumento que permite compreender uma série de aspectos, dentre os quais Guerra (2011) cita a dinâmica da ocupação do espaço urbano, a necessidade de transporte da população, as motivações destes deslocamentos e identificação de padrões de deslocamentos desde um ponto ou zona de origem até um ponto ou zona de destino. Estas zonas podem ser definidas a partir de divisões geográficas (com base em bairros, regionais, etc), via divisões matemáticas (via uso de técnicas de agrupamento de dados), dentre outras.

Ainda de acordo com a Associação Nacional De Transportes Públicos (1997), dentre os métodos comumente adotados para a realização de estudos de OD, tem-se os métodos diretos, ou seja, aqueles que se baseiam na realização de pesquisa de campo. Tais pesquisas

5 datapoa.com.br/group/mobilidade - último acesso em Jun. 06, 2019.
6 dados.prefeitura.sp.gov.br/group/transporte - último acesso em Jun. 06, 2019.
7 dados.natal.br/group/mobilidade-urbana - último acesso em Jun. 06, 2019.
8 dados.recife.pe.gov.br/group/mobilidade - último acesso em Jun. 06, 2019.
9 www.dados.df.gov.br/group/mobilidade - último acesso em Jun. 06, 2019.
10 www.curitiba.pr.gov.br/dadosabertos/ - último acesso em Mar. 17, 2019.
11 ippuc.org.br - último acesso em Mar. 17, 2019.
12 urbs.curitiba.pr.gov.br/ - último acesso em Mar. 17, 2019.

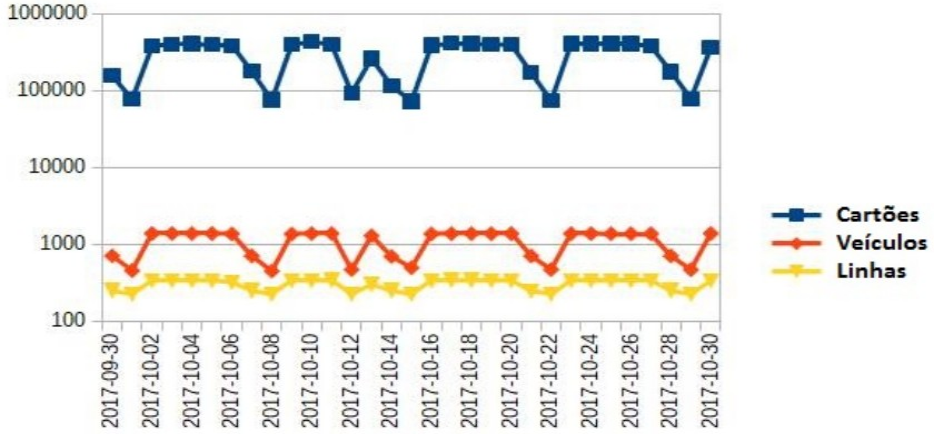
visam a coleta não só da origem e do destino dos deslocamentos, como também das variáveis de tempo a eles associados (início, fim, duração), o modo de transporte utilizado e os motivos da realização do deslocamento, além de informações socioeconômicas. Embora seja uma prática relativamente comum, verifica-se que os empecilhos de realização são cada vez maiores devido às dificuldades de acesso aos entrevistados. Além disso, os custos e o tempo de execução deste tipo de pesquisa costumam ser tão expressivos a ponto de restringir a frequência de sua aplicação para prazos mais alongados. Por outro lado, em cidades onde o Sistema de Transporte Público (STP) faz uso de Sistema de Bilhetagem Eletrônica de Passagens (SBEP), os métodos indiretos podem ser uma alternativa viável aos métodos diretos. Neste caso, em vez da realização de pesquisas de campo, modelos matemáticos de análise de demanda são utilizados para analisar os dados referentes à utilização do STP, proporcionando resultados análogos aos provenientes dos métodos diretos, mas em um tempo mais curto e a um custo menor.

A adoção de SBEP nos veículos que compõem o STP traz consigo uma série de desafios, dentre eles o de gestão de grandes volumes de dados. O STP de Curitiba, por exemplo, possui 9940 paradas de ônibus, 342 estações-tubo, 23 terminais rodoviários e 482 linhas de ônibus que cobrem 9135 vias públicas. Segundo dados do IPPUC[13], o STP de Curitiba transporta, em média, 1,3 milhão de passageiros por dia através de cerca de 15.000 viagens e um total de 300 mil quilômetros percorridos: a Figura 2 traz a quantidade de embarques (registro de uso dos cartões dos usuários), quantidade de veículos em operação e de linhas de ônibus sendo operadas por dia desde 30/09/2017 até 30/10/2017. Dados tais números, não é difícil reconhecer o quão desafiador é gerenciar o volume de dados produzido mensalmente apenas pelo STP de uma cidade como Curitiba.

Neste cenário envolvendo análise de grandes volumes de dados abertos, uma abordagem frequentemente adotada é o uso de diferentes técnicas de visualização de dados, as quais segundo Yu (1977), oferecem uma organização visual dos dados segundo algum critério, possibilitando não só abstrair detalhes do conjunto de informações, mas também evidenciar aspectos talvez não perceptíveis quando analisados os mesmos dados na sua forma bruta. Assim, representações gráficas tradicionalmente empregadas apenas para apresentação de dados passam a ser usadas como ferramentas de exploração de dados. Estudos envolvendo OD podem ser beneficiados a partir da adoção de diferentes técnicas de visualização de dados.

13 www.urbs.curitiba.pr.gov.br/transporte/estatisticas - último acesso em Ago. 26, 2019.

Figura 2 - Números sobre o STP de Curitiba, por dia, desde 30/09/2017 até 20/10/2017.



Fonte: Do próprio autor.

Uma série de trabalhos também tem sido realizados voltados para o estudo de mobilidade urbana sob o ponto de vista de grafos. No caso do STP rodoviário abordado na pesquisa de Silva et al. (2016), os vértices foram utilizados para representar paradas e terminais de ônibus; as arestas para representar as linhas de ônibus que conectavam tais pontos. No trabalho de Chapleau e Morency (2005), um STP multimodal foi representado de forma equivalente, com a variante de que um dado vértice poderia representar uma parada de ônibus, um ponto de metrô e/ou um terminal de trem. Já em Lu et al. (2015), em vez de adotar Pontos de Interesse (*POI*, sigla em inglês) para representar as arestas do grafo, foram utilizadas Regiões de Interesse (*ROI*, sigla em inglês) como forma de aumentar a granularidade dos dados sob análise, passando a restringir os dados sob análise a apenas aqueles contidos em tais perímetros. Adotando tais abordagens, é possível valer-se de conceitos, topologias e métricas da teoria de grafos no estudo de aspectos relacionados a mobilidade urbana contornando, inclusive, desafios relacionados a grandes volume de dados.

De uma perspectiva formal, Añez et al. (1996) menciona que uma rede de transporte pode ser definida como um conjunto de enlaces direcionais da forma $N = (V, L)$, onde V é um conjunto de vértices (nós) $V = \{v_1 , v_2 , ..., v_n \}$ e L é um conjunto de enlaces (arestas) $L = \{l_1 , l_2 , ..., l_n \}$, de tal modo que $L = (v, w, Q_{vw}), v, w \in V$, onde v e w são vértices (nós) de origem e destino, e Q_{vw} é um conjunto de atributos de cada enlace (aresta), tais como distância, capacidade, número de passageiros, velocidade, dentre outros. Como os números de vértices (nós) de origem e de destino são preservados, o modelo pode ser transformado novamente na notação gráfica original, sem nenhum esforço computacional adicional. O mesmo modelo pode ser estendido para redes multidimensionais e multimodais, inclusive. Se considerarmos esta definição no contexto do STP de Curitiba, o conjunto L pode ser apresentado como

mostrado na Tabela 1, onde *v* e *w* poderiam ser pontos de ônibus ou terminais de ônibus, e Q_{vw} um conjunto de parâmetros relevantes para a administração local de ônibus.

Tabela 1 - Exemplo de um conjunto *L*, considerando dados de OD de Curitiba.

v	*w*	Q_{vw}
(-25.460766, -49.345255)	(-25.436778, -49.274341)	("NumLinha=INTERBAIRROS IV", "Veiculo=MC304", "NumCartao=0000558750", "DataHora=2017-10-04 06:20:27", "NataNasc=1987-09-17", "Sexo=F")
(-25.38542, -49.28037)	(-25.417515, -49.246675)	("NumLinha=INTERBAIRR II", "Veiculo=DR113", "NumCartao=0000558806", "DataHora=2017-10-03 07:26:24", "NataNasc=1987-09-16", "Sexo=F")
(-25.422613, -49.300685)	(-25.430181, -49.2724)	("NumLinha=BIGORRILHO", "Veiculo=BC852", "NumCartao=0000559715", "DataHora=2017-10-04 14:38:16", "NataNasc=1982-11-04", "Sexo=F")

Dentre os dados relacionados ao STP de Curitiba, existem componentes geográficos (trajeto de linhas de ônibus, paradas de ônibus, terminais rodoviários, posição dos veículos em operação, etc), componentes temporais (instante do uso do cartão do usuário do STP, instante da coleta da posição dos veículos em operação a cada 05 minutos, momento em que o veículo partiu do início e alcançou o final da linha, dentre outros), além de demais dados (informações dos usuários e de seus respectivos cartões, etc). Para manipulação e processamento destes dados, costuma-se fazer uso de ferramentas como *ArcGIS* e *QuantumGIS*, em que geralmente apenas fatias de dados são analisadas. Análises mais complexas exigem conhecimento de linguagens específicas como *SQL - Structured Query Language*.

Nesse sentido, este trabalho apresenta uma nova ferramenta visual para suportar consultas espaço-temporais sobre os dados de OD. Cada consulta é associada a um conjunto de viagens realizadas no STP de Curitiba, envolvendo dados relacionados tanto à viagem (local e horário de embarque e desembarque, dados da linha e do veículo envolvidos) quanto ao usuário (faixa etária e sexo). Para lidar com a escalabilidade para visualização, também adotou-se um recurso de clusterização semelhante ao utilizado em Vila (2016). As consultas podem ser compostas e disparadas pelo usuário, e tais resultados podem ser exibidos através de diferentes representações visuais e agregados sob diferentes níveis de detalhamento, mantendo sempre os contextos espaciais e temporais.

1.1 Objetivos gerais

Este trabalho apresenta uma nova ferramenta visual para suportar consultas espaço-temporais sobre os dados de OD. Cada consulta é associada a um conjunto de viagens (e

atributos relacionados) baseada em Ferreira et al. (2013), usando a análise de viagens já mencionada em Diniz Junior (2017). Para lidar com a escalabilidade para visualização, fez-se uso de recursos de clusterização de dados georreferenciados Vila (2016).

1.2 Objetivos específicos

Para atingir os objetos gerais, são considerados os seguintes objetivos específicos:

- Realizar uma análise exploratória no conjunto de dados referentes aos registros de utilização dos cartões dos usuários do STP de Curitiba;
- A partir dos registros de utilização dos cartões dos usuários e do posicionamento dos veículos do STP, valer-se da estratégia adotada por Diniz Junior (2017) e buscar estimar o ponto de embarque e de desembarque dos usuários do STP;
- Realizar uma análise exploratória nos dados referentes ao deslocamento a nível intrabairro e interbairro dos usuários do STP de Curitiba;
- Desenvolver uma solução de visualização de OD dos dados geoespaciais abertos da cidade de Curitiba;
- Disponibilizar a solução desenvolvida para que usuários experimentem e a avaliem;
- Analisar os resultados obtidos.

1.3 Metodologia

A metodologia deste trabalho foi dividida em 05 etapas, conforme mostra a Figura 3. A primeira etapa envolveu a elaboração de um referencial teórico relacionado aos objetivos pretendidos. Assim, foram trazidos conceitos de Cidades Inteligentes, de Sistemas de Informação Geográfica - SIG (*GIS*, sigla em inglês), de SIG para Transportes, conceitos e técnicas relacionadas a visualização de dados e conceitos relacionados a teoria dos grafos aplicada ao estudo de STP. Desta forma, pretendeu-se contextualizar a problemática envolvida neste projeto.

A segunda etapa consistiu em buscar trabalhos correlatos que fornecessem subsídios para a realização deste projeto. Nesta direção, dentre uma série de trabalhos analisados, foi dada uma atenção especial para os protótipos de visualização de dados georreferenciados propostos nas pesquisas de Ferreira et al. (2011), Ferreira et al. (2013), Lu et al. (2015), Zhang et al. (2015b) e Vila (2016), buscando identificar em cada um deles quais tecnologias foram utilizadas e como foram combinadas, quais técnicas de visualização de dados foram

empregadas, quais funcionalidades foram disponibilizadas ao usuário e também questões relacionadas ao volume de dados e ao desempenho da aplicação. Também foi dedicada uma atenção especial à metodologia adotada por Diniz Junior (2017) para deduzir pontos de embarque e de desembarques de usuários do STP de Curitiba a partir dos registros de uso dos cartões de usuários (ora chamados de *smart cards*) e de posição dos ônibus do STP de Curitiba. Também foram consideradas as diferentes perspectivas adotadas em IPPUC (2017) para analisar dados referentes a Pesquisa de OD domiciliar realizada em meados de 2017 em Curitiba.

A terceira etapa consistiu na modelagem e desenvolvimento do protótipo de visualização de dados de OD. Para nortear o desenvolvimento, um questionário foi aplicado a um grupo de potenciais usuários, o que nos forneceu subsídios importantes sobre quais funcionalidades a ferramenta deveria oferecer, quais filtros de pesquisa deveriam estar disponíveis e para quais fins a solução seria útil. A partir disso, foi viabilizada a base dados e a interface de usuário do protótipo. Finalmente, o protótipo como um todo recebeu um *tuning* para proporcionar um melhor tempo de resposta das consultas.

Na quarta etapa foram realizados testes do protótipo, tanto por parte dos autores deste projeto quanto por parte de um grupo de usuários, objetivando verificar eventuais falhas e sugestões de melhoria. Na quinta etapa foi realizada uma análise dos resultados obtidos e algumas reflexões sobre os desafios enfrentados e as oportunidades de melhoria identificadas. Nesta etapa foram também trazidas algumas sugestões de trabalhos futuros.

1.4 Publicações

As publicações diretamente relacionadas a este trabalho (em ordem cronológica) são:

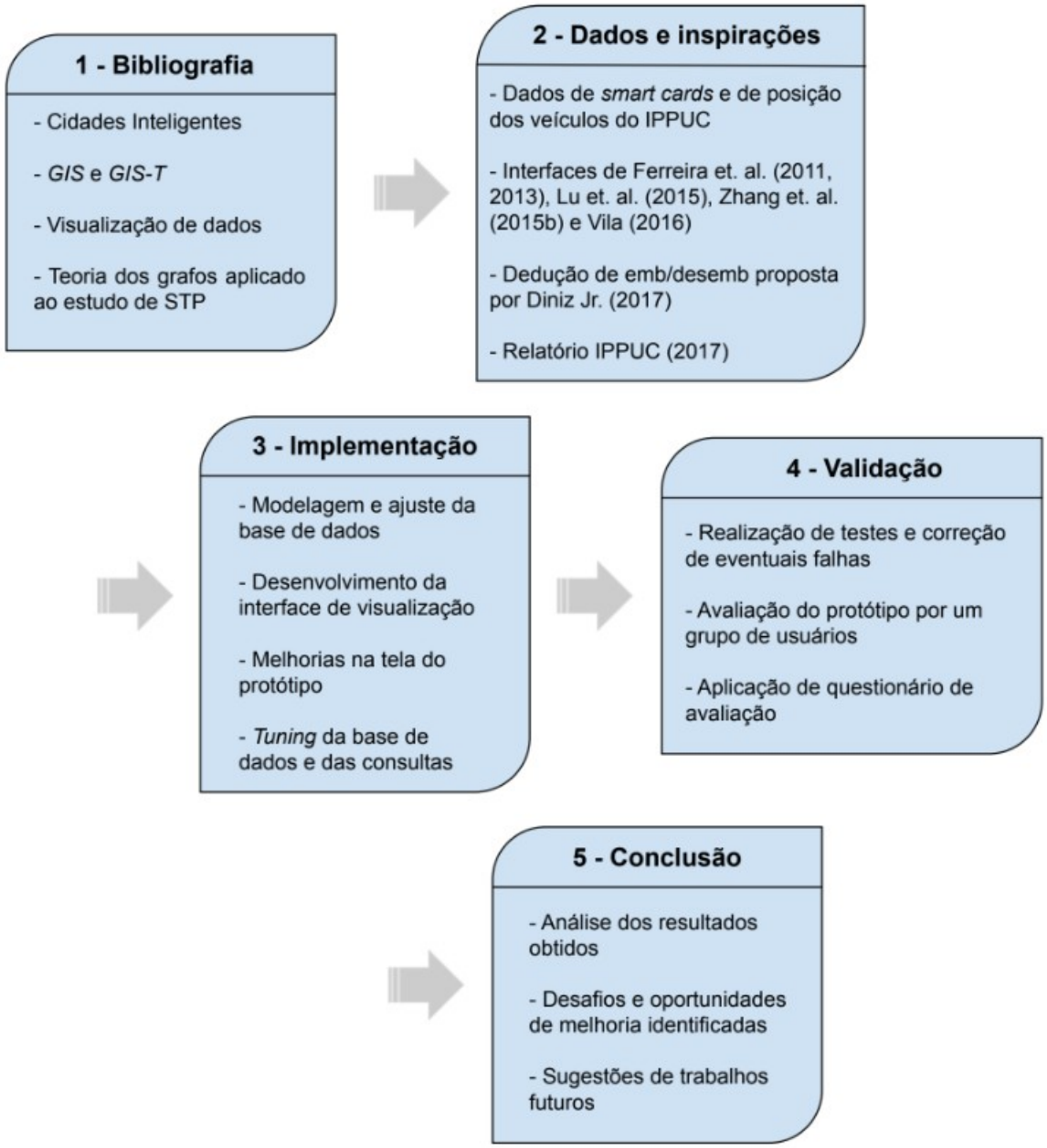
1. PARCIANELLO, Y; KOZIEVITCH, N. P. **Exploratory analysis of public daily expenses from the government of Rio Grande do Sul**. II Workshop de Computação Social. 2017.

2. SIMETTE, G.; PARCIANELLO, Y.; KOZIEVITCH, N. P.; FONSECA, K. V. O. Análise da situação dos redutores de velocidade de Curitiba. In: ESCOLA REGIONAL DE BANCO DE DADOS, 14., Rio Grande do Sul. **Anais** […]. Rio Grande do Sul: SBC, 2018, p. 123–126.

3. KOZIEVITCH, N. P.; PARCIANELLO, Y.; FONSECA, K. V. O.; ROSA, M. O.; GADDA, T. M. C.; MALUCELLI, F. C. Transportation: An overview from Open Data

approach. In: INTERNATIONAL SMART CITIES CONFERENCE, 4., 2018, USA. **Anais** […]. USA: IEEE, 2018, p. 1–8.

Figura 3 - Etapas da metodologia adotada.



Fonte: Do próprio autor.

4. MATTOS, V. G.; VASCONCELOS, P. H.; PARCIANELLO, Y.; KOZIEVITCH, N. P.; BERARDI, R. Visualização dos dados abertos da polícia rodoviária federal sobre acidentes nas rodovias brasileiras. In: SIMPÓSIO BRASILEIRO DE BANCO DE DADOS, 34., 2019, Ceará. **Anais** […]. Ceará: SBC, 2019, p. 193–198.

1.5 Estrutura da dissertação

Este trabalho está organizado da seguinte forma: na seção 2 são apresentados conceitos de Cidades Inteligentes, Sistemas de Informação Geográfica, Sistemas de Informação Geográfica para Transportes, Teoria de Grafos aplicadas ao estudo de Sistemas de Transporte Público, trabalhos correlatos e desafios relacionados à presente pesquisa. Na seção 3 é apresentado o protótipo de visualização de dados de OD, incluindo o processo de levantamento de requisitos, o processo de concepção da base de dados da aplicação, a possibilidade de análise sob diferentes regiões de interesse e, a arquitetura da aplicação. Na seção 4 apresentam-se os resultados do estudo de caso do protótipo aplicado ao STP Curitiba. Por fim, conclui-se o trabalho na seção 5.

2 CONCEITOS BÁSICOS E TRABALHOS RELACIONADOS

Neste capítulo são apresentados conceitos referentes a Cidades Inteligentes, Sistemas de Informação Geográfica com ênfase em visualização de dados, Sistemas de Informação Geográfica para Transportes com ênfase também em visualização, Teoria de Grafos aplicadas ao estudo de Sistemas de Transporte Público, trabalhos relacionados e principais tecnologias utilizadas, além dos desafios relacionados à presente pesquisa.

2.1 Cidades Inteligentes

O conceito de Cidades Inteligentes (*Smart Cities*, em inglês) pode ser entendido como a integração de tecnologias numa abordagem estratégica para promover a sustentabilidade econômica e ambiental, assim como o bem-estar social. Para Aldama-Nalda et al. (2012), uma *Smart City* é aquela que utiliza tecnologias para integrar suas infraestruturas e serviços para melhorar a eficiência, a transparência e a sustentabilidade. É um modelo de cidade no qual convergem diversas correntes de desenvolvimento urbano (cidade sustentável, cidade inovadora, cidade digital, cidade de conhecimento) para melhorar a qualidade de vida e a gestão de recursos e de serviços.

No âmbito legislacional, a norma técnica brasileira voltada para cidades sustentáveis é a NBR ISO37120:2017 - "Desenvolvimento sustentável de comunidades - Indicadores para serviços urbanos e qualidade de vida"[14]. Esta norma é uma tradução e adaptação da norma ISO37120:2014 - "*Sustainable development of communities - Indicators for city services and quality of life*"[15], a qual fora elaborada pela TC-268 (*Technical Committee*)[16]. A NBR ISO37120:2017 propõe uma padronização de indicadores relativos a cidade (serviços urbanos ofertados e qualidade de vida, dentre outros). Desta forma, objetiva-se oportunizar a realização de análises comparativas de diferentes comunidades, análises evolutivas de uma dada comunidade, favorecendo a troca de experiências e de boas práticas Couto (2018).

Para Souza et al. (2015), dentre os assuntos englobados em *Smart Cities*, a mobilidade urbana é um fator crítico para o funcionamento de uma cidade, pois proporciona

14 abntcatalogo.com.br/norma.aspx?ID=366389 - último acesso em Mar. 17, 2019.
15 iso.org/standard/62436.html - último acesso em Mar. 17, 2019.
16 iso.org/committee/656906.html - último acesso em Mar. 17, 2019.

mobilidade aos cidadãos e é através do qual escoam produtos e serviços necessários para o funcionamento das cidades. Neste cenário, o avanço tecnológico tem transformado veículos ordinários em veículos conectados, capazes de transmitir informações em tempo real que podem ser utilizados no (re)planejamento urbano. Tais dados podem fornecer subsídios que permitirão melhorar o desempenho do sistema de transporte (parâmetros operacionais como o consumo de combustível, o tempo de espera e o de deslocamento de usuário, por exemplo), reduzindo os impactos ambientais e sociais do transporte (relacionados à poluição sonora, poluição do ar, engarrafamentos e problemas de saúde decorrentes dos oriundos do trânsito, por exemplo).

2.2 Sistemas de Informação Geográfica

De acordo com Worboys e Duckham (2004), Sistemas de informação geográfica (SIG) são sistemas de informação baseados em computador que proveem recursos para captura, modelagem, manipulação, recuperação, análise e apresentação de dados georreferenciados. Um SIG caracteriza-se como uma tecnologia que utiliza infraestrutura de *hardware* e de *software* para permitir explorar e visualizar informações disponíveis em bancos de dados geográficos.

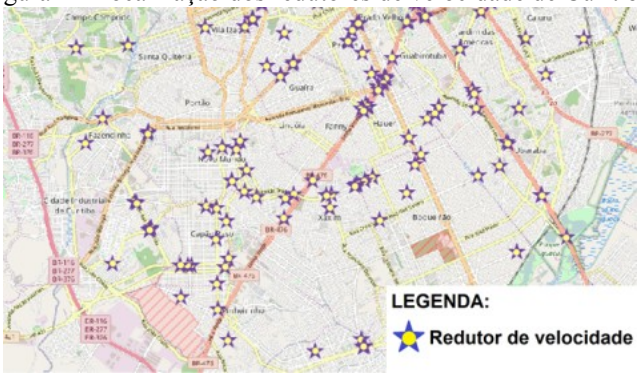
2.2.1 Bancos de dados geográficos

Os SIG distinguem-se dos demais Sistemas de Informação pelos tipos de dados envolvidos, que representam majoritariamente localizações na superfície terrestre em termos de coordenadas geográficas. É neste cenário que ganham ênfase os bancos de dados geográficos que, para Soares (2003), são compostos principalmente por dois grandes módulos: um Sistema Gerenciador de Bancos de Dados (SGBD) e um sistema de processamento de imagens e/ou de dados espaciais.

Segundo Rodrigues (1990), bancos de dados geográficos podem trabalhar com 2 tipos especiais de dados: vetoriais e matriciais. Dados vetoriais (ou dados do tipo vector) são aqueles representados por geometrias (ou elementos geométricos). Neste sentido, uma edificação pode ser representada por uma geometria do tipo ponto, uma estrada por uma do tipo linha, e um bairro por uma do tipo polígono. A lista a seguir aborda cada uma delas.

- **Ponto:** geometria que representa uma única localização no espaço, formada por um par de coordenadas *(x, y)*. Em um mapa de uma cidade, por exemplo, um ponto georreferenciado pode representar pontos de interesse (vide Figura 4);

Figura 4 - Localização dos redutores de velocidade de Curitiba.



Fonte: Simette et al. (2018).

- **Linha:** segmento de linha composta por dois pontos (vide Figura 5);

Figura 5 - Eixos das ruas de Curitiba.

Fonte: Simette et al. (2018).

- **Polígono:** elemento composto por, no mínimo, dois vértices conectados, podendo gerar formas abertas ou fechadas. Neste último caso, acaba por definir duas regiões: uma região interna ao polígono (inscrita) e outra externa ao polígono (circunscrita) (vide Figura 6).

Figura 6 - Contornos dos estados brasileiros.

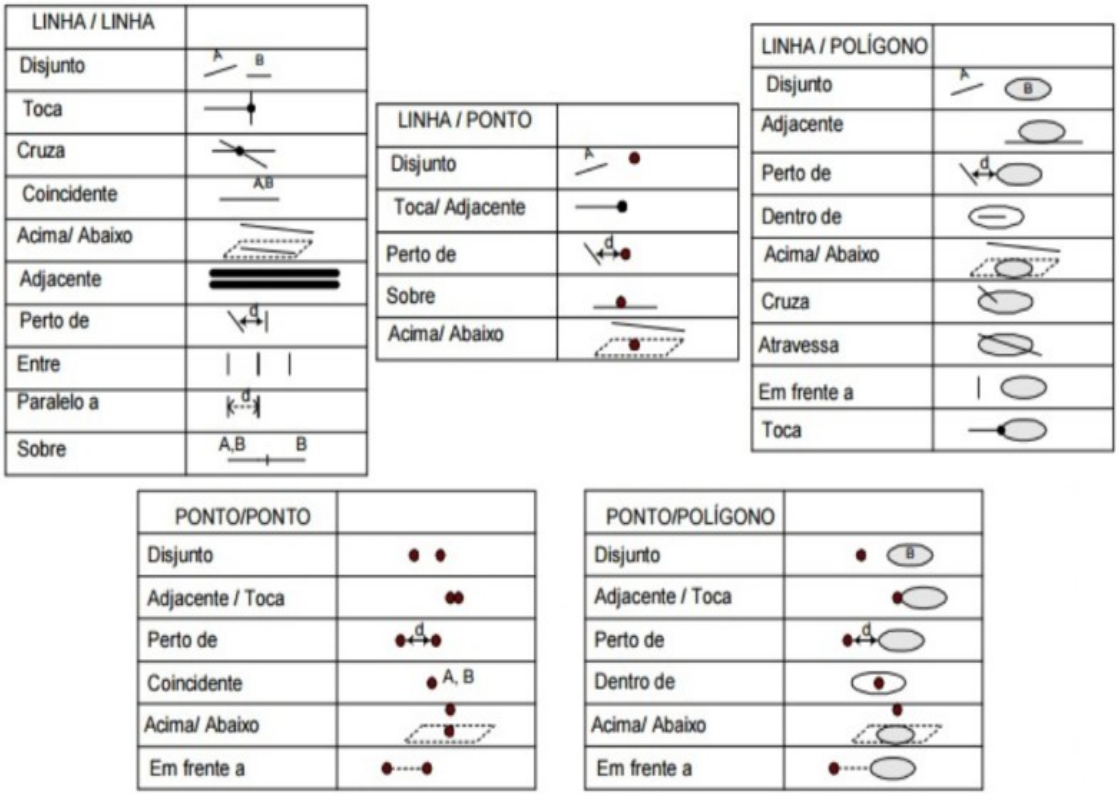
Fonte: Vila (2016).

Ainda, Rodrigues (1990) define dados matriciais (ou imagens matriciais, ou dados do tipo *raster*) como sendo uma matriz de pontos, ou seja, dados compostos por *pixels*. Assim, uma característica básica de um dado matricial é a sua resolução espacial, ou seja, a nitidez deste dado. A Figura 7 traz um exemplo de dado matricial georreferenciado: uma foto da superfície terrestre capturada via satélite. Percebe-se que a busca por uma boa resolução é crucial para permitir estudos mais sofisticados e precisos. Já a Figura 8 traz alguns exemplos de possíveis relações espaciais entre objetos geográficos.

Figura 7 - Explorando um dado rasterizado via satélite.

Fonte: do próprio autor.

Figura 8 - Possíveis relações topográficas entre objetos geográficos.



Fonte: Kono (2016).

2.2.2 Visualização de Dados no âmbito de SIG

Para Card (1999), a visualização é mais do que uma representação gráfica de dados ou conceitos: significa construir uma imagem visual na mente humana. Já para Ramos (2005), é o processo de realizar representações visuais concretas para tornar contextos e problemas espaciais visíveis, engajando-se às mais poderosas habilidades humanas para o processamento de informação, aquelas associadas à visão. Dadas tais definições, a seguir são abordadas algumas possibilidades de visualização de dados.

Mapas de fluxo (*Flow Maps*): De acordo com *The Data Visualization Catalogue*[17], tais mapas mostram o movimento de informações ou objetos de um local para outro e sua quantidade. A Figura 9 mostra um mapa de fluxo onde é possível visualizar as origens e os destinos dos deslocamentos dos servidores vinculados à Segurança Pública que participaram da Operação Litoral-RS durante os anos de 2004 a 2017.

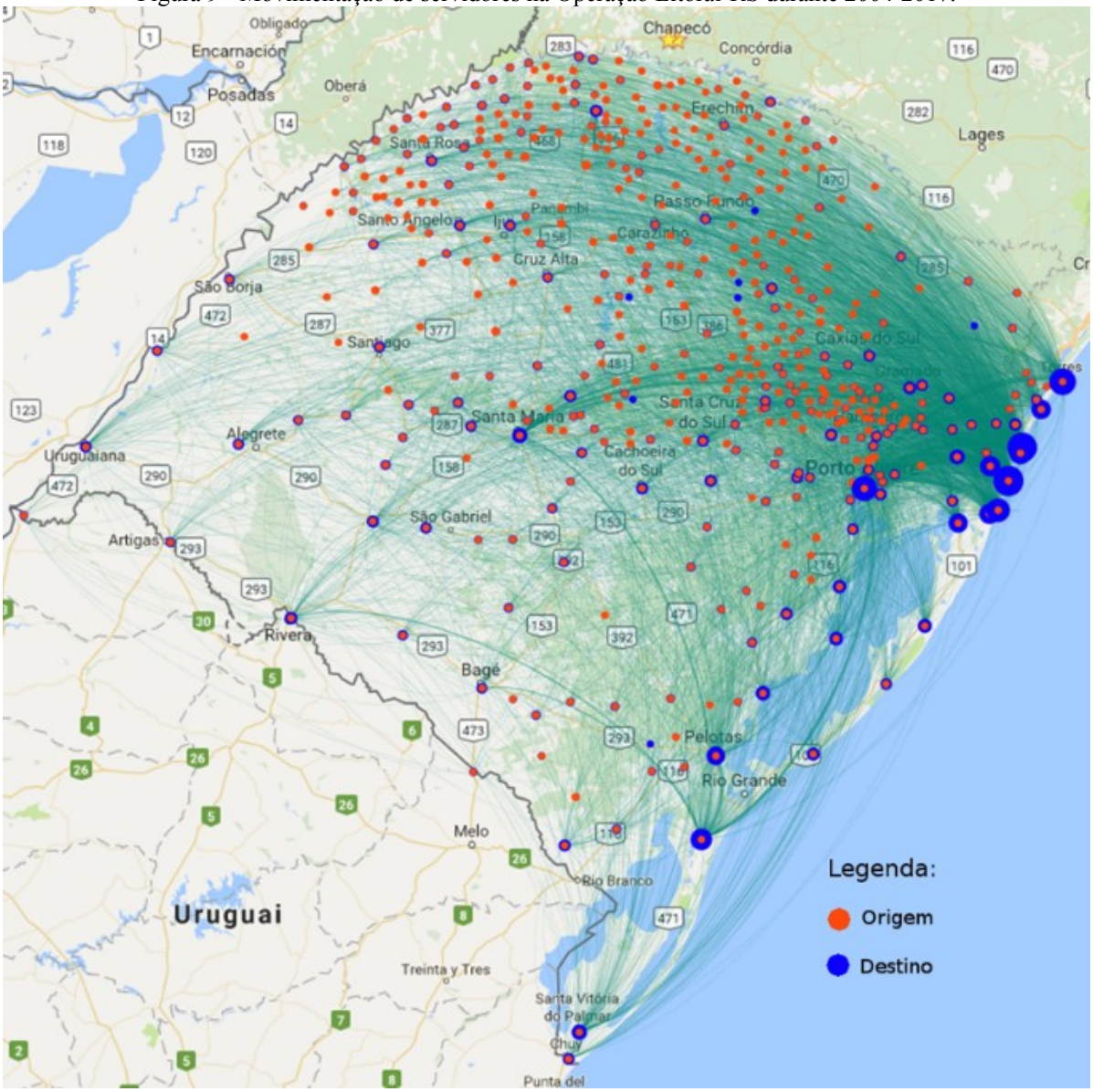
Diagramas de cordas (*Chord Diagrams*): De acordo com *The Data Visualization Catalogue*[18], diagramas deste tipo mostram as relações entre entidades. As conexões entre as entidades (as cordas) são usadas para mostrar que elas compartilham algo em comum. A Figura 10 mostra um diagrama de cordas que representa a situação da migração internacional de pessoas em 2015: a largura das cordas indicam a quantidade de pessoas em situação de migração.

clusterização de marcadores (*Marker clustering*): Zaiane et al. (2002) descreve o processo de clusterização de dados como um agrupamento de informações, considerando: (i) existência de uma forte similaridade entre os elementos pertencentes ao mesmo grupo; (ii) existência de uma fraca similaridade de elementos pertencentes a grupos diferentes. Assim, clusterização de dados pode ser entendida como um processo de abstração de dados a partir do agrupamento de *Markers* (marcadores que representam a localização geográfica de um determinado elemento) para a criação dos *clusters* (os agrupamentos de *Markers*). A Figura 11 trata do número de licenças de negócio concedidas pela prefeitura municipal de Curitiba para abertura de estabelecimentos comerciais nos bairros Batel e Centro daquele município: na Figura 11A, é possível visualizar o total de todas as licenças de negócio emitidas desde 1980

17 datavizcatalogue.com/*Method*s/flow_map.html - último acesso em Mar. 17, 2019.
18 datavizcatalogue.com/*Method*s/chord_diagram.html - último acesso em Mar. 17, 2019.

até 2015. Já a Figura 11B permite visualizar o total de licenças para instalação e operação de lancherias (azul) e restaurantes (amarelo) durante o mesmo período.

Figura 9 - Movimentação de servidores na Operação Litoral-RS durante 2004-2017.



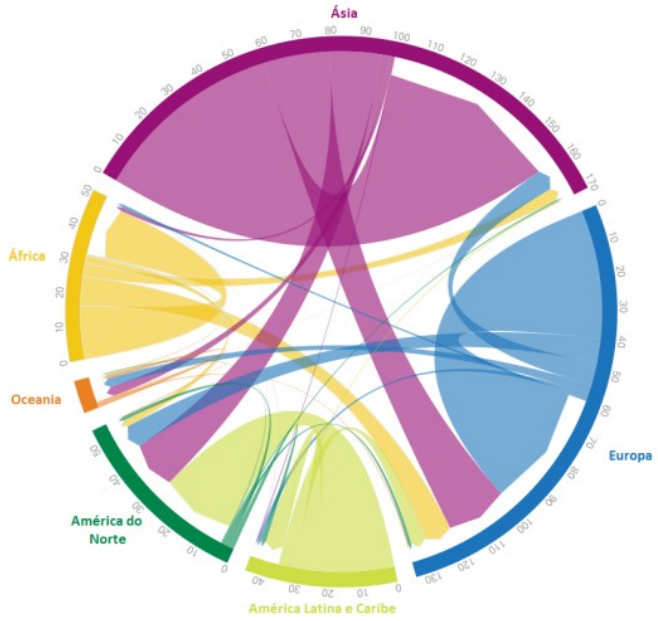
Fonte: Parcianello e Kozievitch (2017).

Mapas de calor (*Heat Maps*): Segundo *GMaps API*[19], *Heat Map* é uma visualização que demonstra a intensidade (frequência de ocorrência) dos dados em regiões do mapa. Em mapas de calor, é comum ser possível definir o gradiente de cores e o raio de influência de cada ponto de dados. A Figura 12 trata dos destinos das viagens oficiais de servidores públicos do Rio Grande do Sul subsidiadas por diárias públicas pagas pelo governo gaúcho: percebe-se na Figura 12B que a região de Porto Alegre, Caxias do Sul e arredores são os

19 developers.google.com/maps/documentation/*Javascript*/heatmaplayer – último acesso em Mai. 31, 2018.

principais destinos de viagens oficiais subsidiadas por diárias públicas do RS. Tal fato não fica tão evidente na Figura 12A.

Figura 10 - Quantitativo de pessoas em migração internacional em 2015.

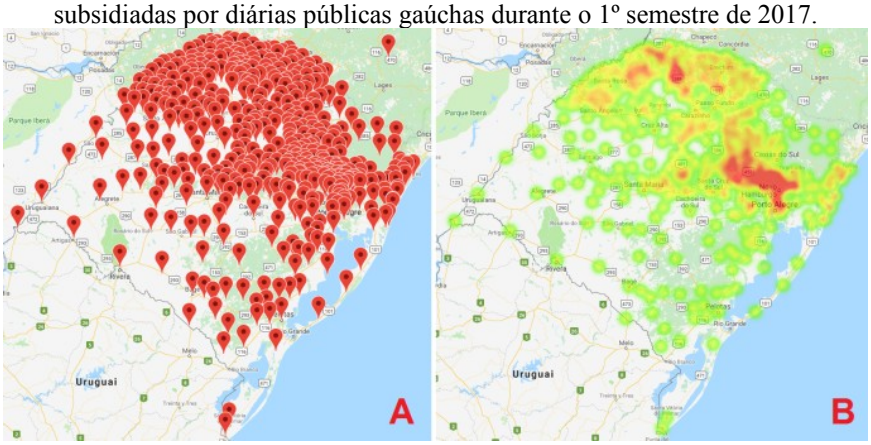


Fonte: Adaptado de Unicef (2017).

Figura 11 - Total de licenças de negócio emitidas desde 1980 até 2015 para os bairros Batel e Centro de Curitiba (A), e total de licenças de negócio para bares e restaurantes nos mesmos bairros (B).

Fonte: Kozievitch et al. (2017).

Figura 12 - Visão de municípios (A) e visão termal (B) dos destinos de viagens subsidiadas por diárias públicas gaúchas durante o 1º semestre de 2017.



Fonte: Parcianello e Kozievitch (2017).

2.3 Sistemas de Informação Geográfica para Transportes

De acordo com Rodrigue (2017), um Sistema de Informação Geográfico para Transportes - SIGT (*GIST*, sigla em inglês) possui quatro componentes básicos que são:

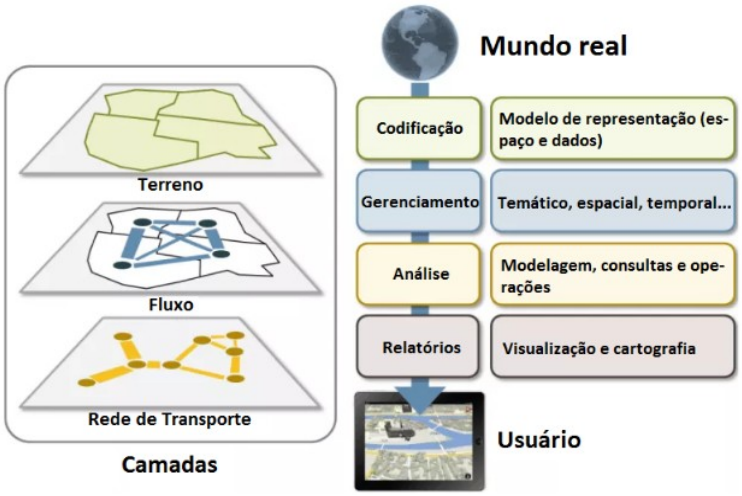
- **Codificação:** Aborda questões relativas à representação de um sistema de transporte e seus componentes espaciais. Para ser útil em um *GIS*, uma rede de transporte deve ser codificada corretamente, implicando em uma topologia funcional composta de nós e enlaces;

- **Gerenciamento:** As informações codificadas são frequentemente armazenadas em um banco de dados e podem ser organizadas a partir de diferentes abordagens: por divisão política (por região, país, unidades censitárias, etc.), por temática (por rodovia, trânsito, ferrovia, terminais etc.) ou por tempo (por ano, mês, semana, etc.);

- **Análise:** Considera a gama de metodologias e ferramentas disponíveis para questões de transporte;

- **Relatórios:** As informações em um SIG são frequentemente organizadas na forma de camadas, sendo estas um conjunto de características geográficas ligadas aos seus atributos. Na Figura 13, pode ser vista a ilustração de um sistema de transporte representado em três camadas: terreno, fluxo (interações espaciais) e rede de transporte. Cada camada tem seus próprios recursos, dados relacionados e pode ser usada de forma individual ou combinada com outras camadas.

Figura 13 - Componentes básicos de um SIG.

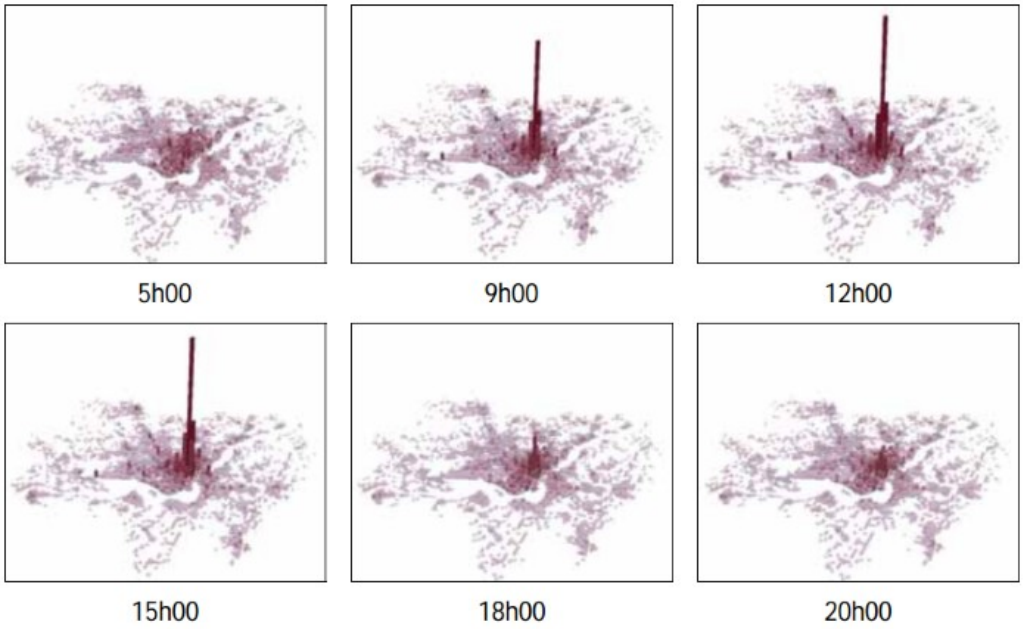


Fonte: Adaptado de Rodrigue (2017).

2.3.1 Visualização de dados no âmbito de SIG-T

A pesquisa de Chapleau e Morency (2005) permite visualizar a influência de um STP na dinâmica da ocupação do espaço urbano ao longo do dia. A Figura 14 indica a variação da densidade demográfica de Montreal com base nos embarques e desembarques de usuários do STP daquela cidade. Outro trabalho análogo pode ser visto em *Manhattan Population Explorer - MPE*[20]. Neste caso, a solução permite explorar a variação da ocupação do espaço urbano da ilha de Manhattan (vide Figura 15).

Figura 14 - Análise espaço-temportal da variação demográfica de Montreal.



Fonte: Chapleau e Morency (2005).

2.3.2 Algoritmos de clusterização

Outra técnica que pode ser utilizada para analisar STP é a utilização de algoritmos de clusterização. Tal escolha se dá quando o nível inicial de detalhamento dos dados não satisfaz os objetivos pretendidos. Neste sentido, busca-se atingir um nível mais alto de abstração a partir da utilização de algoritmos de clusterização.

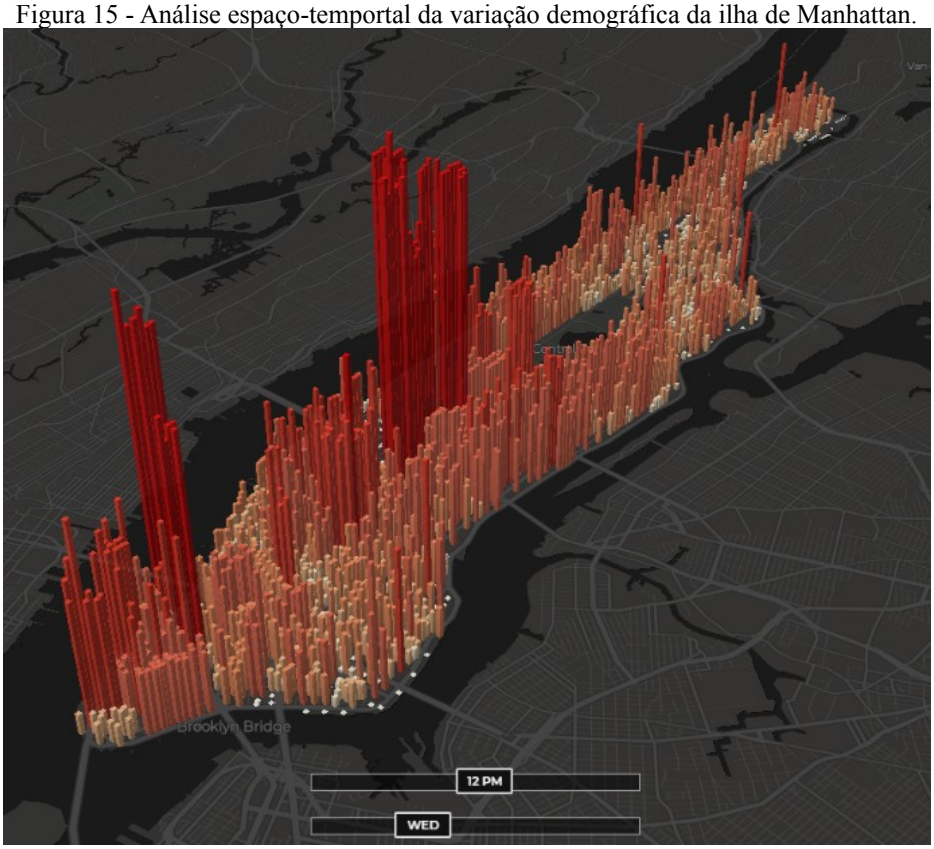
De acordo com Cassiano (2014), a clusterização de dados (ou análise de agrupamentos) é uma técnica de mineração de dados multivariados que tem por objetivo

20 http://manpopex.us/ - último acesso em Nov. 11, 2019.

agrupar os *n* casos da base de dados em *k* grupos denominados *clusters* (ou agrupamentos). Na Literatura, a clusterização de dados pode também ser chamada de análise de *clusters*, *clustering*, *Q-analysis, Typology, Classification Analysis* ou *Numerical Taxonomy*.

Existe uma série de algoritmos de clusterização de dados, e o *k-means* (ou k-médias) é um deles. Dado um *dataset* composto por *n* registros, o *k-means* particiona tais registros (também chamados de ocorrências) em *k* subconjuntos (também chamados de classes ou *clusters*).

Figura 15 - Análise espaço-temportal da variação demográfica da ilha de Manhattan.



Fonte: Saberi (2017).

Para Cole (1998), toda clusterização é feita com objetivo de maximizar a homogeneidade dentro de cada *cluster* e maximizar a heterogeneidade entre diferentes *clusters*. Neste sentido, o autor propõe a seguinte definição para o problema de clusterização: dado um conjunto de *n* elementos $X = \{X_1 , X_2 , ..., X_n\}$, tais elementos serão agrupados em *k* diferentes *clusters* $C = \{C_1 , C_2 , ..., C\,k\}$ de modo que $C_1 \cup C_2 \cup ... \cup C_n = X, C_i \neq \emptyset$ e $C_i \cup C_j$ para todo i , j.

O processo de clusterização de dados ocorre com base nas semelhanças calculadas a partir dos dados sob análise. Conforme Hartigan (1975), um método utilizado para medir similaridade dos dados é através do cálculo de distâncias entre pares de objetos. Neste

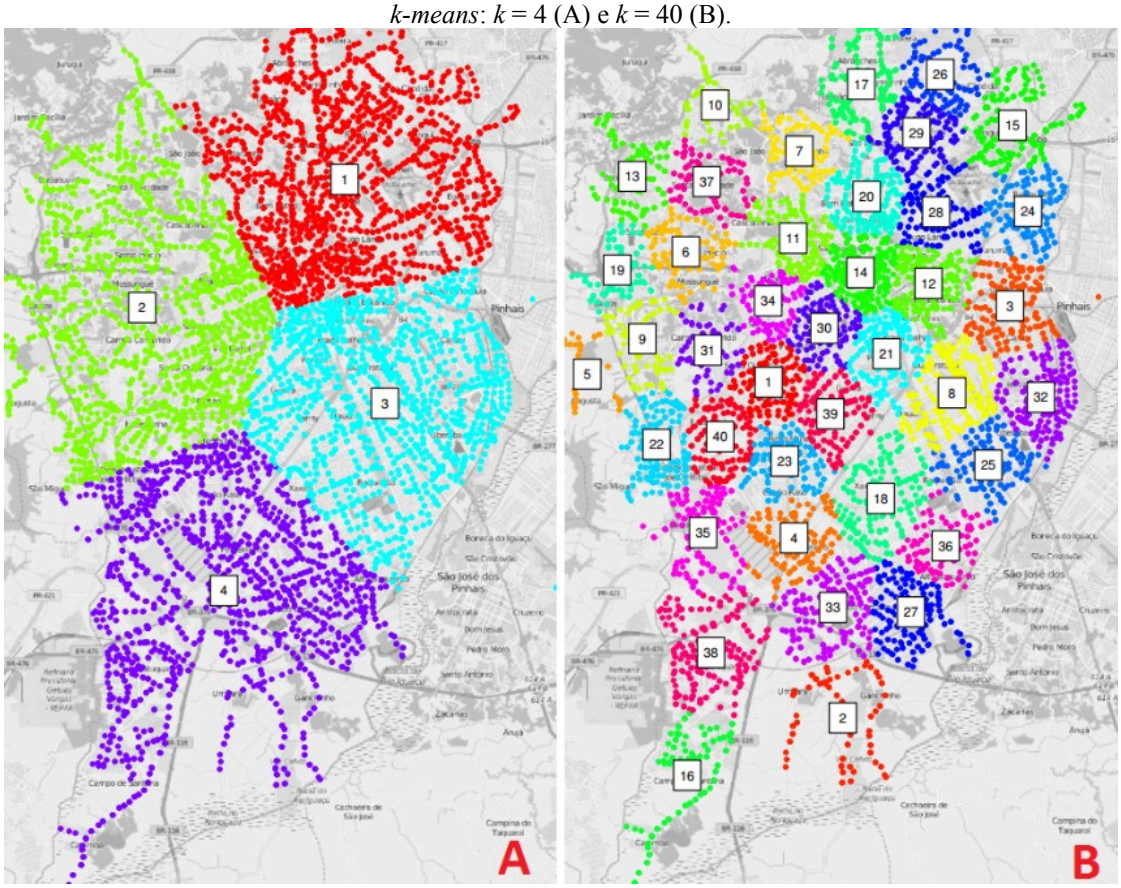
sentido, o autor apresenta alguns dos métodos de cálculo de distância entre objetos utilizados por algoritmos de clusterização, tais como: Distância Euclidiana (Equação 1), *City-Block* (Equação 2) e *Minkowski* (Equação 3). A Figura 16 mostra os resultados das clusterizações das paradas de ônibus de Curitiba obtidas através do algoritmo *k-means*.

$$d(X_i, X_j) = \sqrt{(X_i - X_j)'(X_i - X_j)} = [\sum_{l=1}^{p} (x_{il} - x_{jl})^2]^{\frac{1}{2}} \quad (1)$$

$$d(X_i, X_j) = \sum_{l=1}^{p} |x_{il} - x_{jl}| \quad (2)$$

$$d(X_i, X_j) = [\sum_{l=1}^{p} (x_{il} - x_{jl})^m]^{\frac{1}{m}} \quad (3)$$

Figura 16 - Clusterização das paradas de ônibus de Curitiba obtidas através da utilização do algoritmo *k-means*: $k = 4$ (A) e $k = 40$ (B).

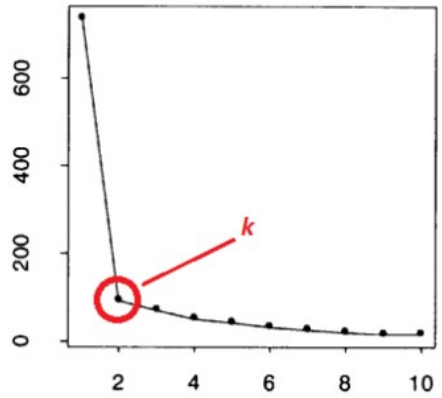


Fonte: Silva et al. (2016).

Com relação às classes, Stolfi et al. (2017) citam o método de *Elbow* como um dos mais utilizados para determinar um número ótimo de *clusters* para particionar um *dataset*. Assim, a partir da variância dos dados em relação ao número de *clusters*, é considerado um valor ideal de *k* quando o aumento no número de *clusters* não representar um valor

significativo de ganho. Observando o gráfico da Figura 17, isto percebe-se o cotovelo (o *Elbow*) quando o número de *clusters* é igual a 2: logo, este é o valor ótimo para *k*.

Figura 17 - Gráfico obtido a partir da utilização do método de *Elbow* para obtenção de um valor ótimo para *k*.



Fonte: Tibshirani et al. (2001).

2.4 Teoria dos grafos aplicados ao estudo de STP

Diversas iniciativas têm sido realizadas no sentido de estudar Sistemas de Transportes Públicos sob o ponto de vista de grafos. No caso do STP rodoviário abordado na pesquisa de Silva et al. (2016), os vértices foram utilizados para representar as paradas e os terminais de ônibus; as arestas para representar as linhas de ônibus que conectavam tais pontos. No trabalho de Chapleau e Morency (2005), um STP multimodal foi representado de forma equivalente, com a variante de que um dado vértice poderia representar uma parada de ônibus, um ponto de metrô e/ou um terminal de trem.

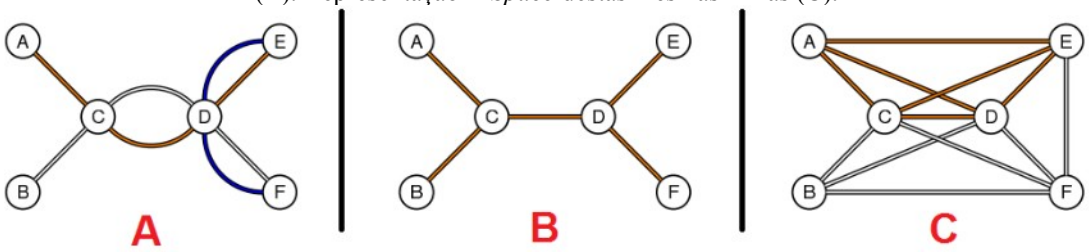
2.4.1 Topologias

De acordo com Zhang et al. (2015a), a aplicação da teoria dos grafos ao estudo de STP pode ser dada a partir de diferentes abordagens. Como exemplos, podemos citar a representação direta, a topologia *L-Space* e a topologia *P-Space*. Para explicar a primeira delas, tomemos como base o grafo da Figura 18A: nela são representadas 03 linhas hipotéticas de ônibus de um STP: a linha Laranja (que passa pelos pontos A, C, D e E), a linha Branca (que passa pelos pontos B, C, D e F) e a linha Azul (que passa pelos pontos D, E e F). Cada nó representa uma parada de ônibus e cada enlace representa uma conexão, caso exista uma

linha de ônibus que vá de um ponto até outro. Percebe-se alguns pontos em comum e alguns enlaces múltiplos entre pares de nós.

Conforme Von Ferber et al. (2009), uma representação *L-Space* deste grafo pode ser vista na Figura 18B. A diferença principal entre estas representações é a quantidade de enlaces entre os nós: Embora possam existir diferentes formas de ir de C para D, ou de D para E, ou de D para F, na topologia *L-Space* nenhum enlace múltiplo é permitido. Já na topologia *P-Space*, conforme abordado por Von Ferber et al. (2009) e por Zhang et al. (2015a), representa-se através de arestas não apenas as conexões entre pontos consecutivos das linhas do STP: se uma linha de ônibus contém vários pontos, então todos estes pontos estão conectados (na linha Laranja, de A se pode ir até C, D e/ou E, por exemplo) e cada uma destas conexões deve ser representada no grafo através de um enlace. Desta forma, a Figura 18C traz a representação *P-Space* deste STP hipotético: em destaque na cor laranja está o subgrafo *P-Space* referente a linha Laranja mostrada na Figura 18A.

Figura 18 - Mapa de linhas hipotéticas de ônibus de um STP (A). Representação *L-Space* destas linhas (B). Representação *P-Space* destas mesmas linhas (C).



Fonte: Adaptado de Von Ferber et al. (2009).

A teoria dos grafos oferece métricas que podem ser aplicadas ao estudo de mobilidade urbana. A centralidade de grau, conforme abordado em Borba (2013), pode indicar o número de linhas de ônibus que passam por determinado ponto de ônibus. Já a centralidade de proximidade, segundo Sabidussi (1966), pode ser utilizado para verificar o quão central é determinado ponto de ônibus em relação a todos os demais pertencentes a rede de transporte. Finalmente, Wasserman e Faust (1994) sugere o uso da centralidade de intermediação como um indicador de número de linhas de ônibus que passam por determinado ponto.

2.5 Trabalhos correlatos

Percebe-se que existe um expressivo número de produções científicas que envolvem questões relativas às cidades inteligentes, utilização e integração de dados abertos e sistemas

de informação geográficos. A seguir são abordados alguns destes trabalhos, os quais contribuíram de diferentes maneiras para a realização da presente pesquisa.

2.5.1 Análise de dados abertos

Em Beluzo (2015), foram integrados dados oriundos de diferentes portais governamentais de dados abertos que tratavam de receitas e despesas de municípios do estado de São Paulo. Durante sua pesquisa, foram identificados problemas de qualidade dos dados, problemas de incompatibilidade de dados oriundos de diferentes fontes e problemas relacionadas aos metadados de alguns *dataset*s utilizados.

Jerônimo et al. (2016) analisaram a correlação entre dados de *tweets* (dos usuários do *Twitter*) e de indicadores socioeconômicos de Londres. Como resultados, foram identificaram significativas correlações entre padrões de mobilidade e indicadores socieconômicos como empregabilidade e qualificação profissional. A pesquisa ressalta a relevância das redes sociais virtuais como provedores de dados no âmbito de *Open Data*.

Na pesquisa de Dias et al. (2016) foi realizado um estudo comparativo de dados econômicos, demográficos e de frota de uma série de cidades brasileiras. Então, tais cidades foram organizadas em *clusters* através do uso do algoritmo *k-means*. Como resultados, os *clusters* obtidos poderiam subsidiar a criação de grupos de trabalho intermunicipais no sentido de juntos comporem seus Planos Municipais de Mobilidade Urbana.

Na pesquisa de Osama et al. (2015), foi analisada a qualidade do ar de alguns países europeus. Os dados utilizados foram coletados por estações de monitoramento localizadas nos diferentes países sob estudo. Então, utilizou-se o algoritmo *k-means* para identificar regiões onde o ar é menos poluído e outras onde o ar é mais insalubre.

2.5.2 Mobilidade urbana

O trabalho de Barczyszyn (2015) evidencia relevância da análise exploratória de dados abertos georreferenciados no (re)planejamento urbano de uma cidade como a de Curitiba. Na pesquisa são abordadas questões relacionadas a integridade de dados e contrastes relacionados ao formato e nível de detalhamento dos dados.

Em Kozievitch et al. (2017), é realizada uma análise espaço-temporal georreferenciada das atividades de negócio de Curitiba durante um período de 30 anos. O trabalho, no qual foram integrados dados abertos oriundos de diferentes fontes, permite compreender como seu deu o desenvolvimento da cidade ao longo do período analisado e o impacto da malha viária e do STP na ocupação do espaço urbano daquela cidade.

No trabalho de Vila (2016), é proposta uma solução *web-mobile* que permite a visualização espaço-temporal de dados georreferenciados. Para a apresentação dos resultados, foram utilizados recursos gráficos como marcadores, mapa térmico e clusterização de marcadores. A análise das avaliações dos usuários indicou o quão assertiva foi não só as possibilidades de filtragem de dados como também as diferentes possibilidades de visualização dos resultados oferecidos pela solução.

No trabalho de Gay et al. (2017), foram mapeados os locais mais vulneráveis da cidade de São Paulo quanto a possibilidade de inundação e nível de acessibilidade da região. Tais feitos foram alcançados a partir de dados abertos georreferenciados da hidrografia, das vias de acesso e da altimetria da região. Os resultados reforçam a relevância de estudos desta natureza em termos de mobilidade urbana e de defesa civil.

Na pesquisa de Vila et al. (2016), são analisados dados abertos de que tratam do transporte público de Curitiba. A pesquisa envolveu a realização de análise exploratória de dados abertos, integração de dados, utilização de algoritmos de clusterização (*k-means* e *Marker cluster*er) e teoria de grafos no estudo do STP daquela cidade.

A pesquisa de Saberi et al. (2017) traz uma análise comparativa dos STP das cidades de Chicago e Melbourne utilizando modelos de rede complexas. A abordagem adotada visa facilitar a compreensão da estrutura, das interações e evolução da demanda das viagens naquelas cidades. No artigo, sugere-se que os processos subjacentes à demanda de viagens, vistos como uma rede, também sejam impulsionado pela força da interação entre locais (ou nós).

Em Costa et al. (2017), são analisados os dados abertos que tratam principalmente de redutores de velocidade da cidade de Curitiba. Neste sentido, foram integrados dados de radares, lombadas, linhas de ônibus e divisão de bairros daquela cidade. A pesquisa identificou uma série de problemas da ordem de integração de dados abertos.

Na pesquisa de Simette et al. (2018), os dados de localização dos redutores de velocidade de Curitiba foram integrados com dados da malha viária daquela cidade. Tais dados foram então confrontados com as legislações que norteiam o emprego de redutores de

velocidade. Os resultados obtidos indicaram indícios de que uma série de dispositivos já instalados nas vias públicas daquela cidade estariam posicionados em locais impróprios.

Na pesquisa de Kozievitch et al. (2018), é realizado um estudo comparativo baseado nos dados abertos do STP de Curitiba com o de Nova Iorque. Tal estudo permitiu identificar similaridades e contrastes entre os STP existentes em cada uma das cidades analisadas e evidenciou a importância de se buscar implementar um sistema inter e multimodal de transporte público. Os resultados foram obtidos via uso de *PostgreSQL*[21] e *PostGIS*[22].

A pesquisa de Spadon et al. (2018) também compara dados abertos da rede viária, da demografia, da extensão territorial e de outros indicadores urbanos de 645 cidades do estado de São Paulo. O estudo também aplica conceitos de redes complexas e algoritmos de clusterização para classificar tais cidades por semelhança sob diferentes perspectivas.

A pesquisa de Cruz et al. (2018) propõe a identificação e a classificação de anomalias no sistema de transporte rodoviário urbano do Rio de Janeiro através de um processo que envolve integração de *Open Data* e mineração de dados via uso do algoritmo *Apriori*.

Em Silva et al. (2016), utilizou-se o algoritmo *k-means* para analisar a distribuição das paradas de ônibus de Curitiba. Assim, foram identificadas regiões mais bem servidas pelo STP e outras nem tanto. Estudos como este fornecem subsídios para que administração pública repense a distribuição de paradas de ônibus na cidade, de maneira a servir a população de todas as diferentes regiões de uma forma mais igualitária.

No trabalho de Andrade et al. (2014), visando reduzir o tempo de deslocamento de um grupo de pessoas desde as suas respectivas casas até o Campus II do CEFET-MG, os pesquisadores utilizaram o algoritmo *DBScan* para clusterizar os deslocamentos destas pessoas. Assim, as pessoas se deslocariam a pé para o ponto mais próximo (o centróide do *cluster*), onde então embarcariam em veículos compartilhados para deslocarem-se até o destino. Assim, foi possível reduzir uma média de 40% do tempo de deslocamento dos participantes.

A pesquisa de Minetto et al. (2016) utiliza os dados abertos de Curitiba para propor um planejador de rotas para cadeirantes. Para tanto, são utilizados algoritmos de caminho mínimo para propor rotas que considerem calçadas, faixas de pedestres, rampas de acessibilidade e altimetria do percurso. Ainda, a solução proposta faz uso do conceito de *crowdsourcing*, permitindo que o usuário avalie a(s) rota(s) sugerida(s).

---

21 www.postgre*SQL*.org/download/linux/ubuntu/ - último acesso em Mai. 01, 2019.
22 *PostGIS*.net/2013/11/08/*PostGIS*-2.1.1/ - último acesso em Mai. 01, 2019.

2.5.3 Problemas de OD

Em Diniz Junior (2017), é proposto um método para estimar o ponto de embarque e o de desembarque de usuários do STP de Curitiba a partir dos dados de utilização de cartões e nos de geolocalização dos veículos em operação. Na pesquisa de Li Weigang et al. (2002), é proposta uma técnica voltada para estimar o tempo de chegada dos veículos nas paradas de ônibus baseando-se, para tanto, no traçado do linha de ônibus, na posição atual do veículo, na velocidade média atual do veículo e na distância deste até o destino em questão.

No trabalho de Parcianello e Kozievitch (2017), foram analisados os gastos com diárias públicas realizados pelo governo gaúcho durante os anos de 2004 a 2017. No estudo, foram utilizadas técnicas de análise exploratória, aplicação do conceito de grafos no estudo de OD e utilização de diferentes formas de visualização gráfica de dados georreferenciados.

Na pesquisa de Lu et al. (2015), é proposta uma solução que, através do uso da ferramenta *KronoMiner*[23], permite explorar dados de viagens de taxi de Pequim. Na solução, o usuário informa uma *ROI* (*Region of Interest* ou região de interesse) de origem, outra de destino e um dado intervalo de tempo. Cada *ROI* é definida através de um círculo pode ser movido e redimensionado sobre um mapa. A solução então retorna as viagens que satisfazem os critérios de busca.

Nesta mesma direção, a solução proposta por Zhang et al. (2015b) permite que o usuário defina as *ROI* através não apenas da manipulação de 02 círculos, mas desenhando, movimentando e redimensionando qualquer tipo de polígono fechado sobre um mapa. Soluções envolvendo uso de *ROI* otimizam o tempo de consultas, visto que o volume de dados envolvidos se restringem àqueles contidos nas *ROI* envolvidos.

Na pesquisa de Ferreira et al. (2013) foi proposto o *Taxivis*[24], solução desenvolvida para usuários leigos, através da qual é possível pesquisar em um mapa de viagens de táxi de Nova Iorque cujos extremos estejam localizados dentro de diferentes *ROIs*. Tais *ROIs* podem ser manipuladas pelo próprio usuário e os resultados fornecidos pela solução são dados representados por diferentes formas gráficas. Uma pesquisa semelhante é vista em Lu et al. (2015), em que é apresentado um protótipo cuja visualização dos dados de OD se dá na forma de um gráfico circular interativo.

23 http://www.cs.toronto.edu/~fchevali/resources/projects/kronominer/ - último acesso em Set. 07, 2019.
24 github.com/ViDA-NYU/TaxiVis - último acesso em Mar. 24, 2019.

Soluções que se propõem a apoiar a análise de dados de OD podem ser aplicáveis a diferentes contextos. A pesquisa de Ferreira et al. (2011), por exemplo, é apresentada a solução *BirdVis*[25], uma aplicação que permite explorar dados não de mobilidade urbana, mas sim da movimentação migratória de aves. A solução combina interessantes recursos, dentre os quais cita-se a possibilidade de adição de múltiplos mapas, lado a lado, para visualização simultânea e comparação de diferentes espécies de pássaros, seus hábitos alimentares, suas possíveis regiões de origem e prováveis regiões de destino.

2.5.4 Apresentação de dados

O projeto *DataViz*[26] pode ser visto como uma espécie de glossário de possibilidades de visualizações de dados. Através do sítio web do projeto é possível buscar possibilidades de visualização de dados por tipo, por estrutura dos dados de entrada, por função pretendida e por formato da visualização. Outro projeto que segue esta linha é o *City Geographics*[27], que disponibiliza não só diferentes cases de recursos interativos de visualização de dados abertos como também publicações e ideias de pesquisa relacionadas à forma, dinâmica e sustentabilidade urbanas.

Também foram analisadas soluções que visam a disponibilizar dados abertos não apenas em formato *machine readable*, mas também de forma interativa *online*, possibilitando a exploração e análise de dados de forma visual. Um exemplo é a solução disponível em *Manhattan Population Explorer*[28], a qual permite visualizar a dinâmica da ocupação do espaço urbano da ilha de Manhattan. Outro exemplo é o *dashboard* de acidentes de trânsito desenvolvido pelo IPPUC[29], que oferece diferentes filtros e exibe os resultados através de diferentes formatos gráficos.

2.5.5 Principais tecnologias utilizadas nos trabalhos correlatos

Revisões de conceitos, exemplos de diferentes metodologias, uma série de possibilidades e de desafios, enfim, inúmeras foram as contribuições dos trabalhos correlatos

25 http://www.birdvis.org/ - último acesso em Set. 07, 2019.
26 datavizproject.com - último acesso em Mar. 24, 2019.
27 citygeographics.org - último acesso em Abr. 11, 2019.
28 manpopex.us - último acesso em Mar. 24, 2019.
29 ippuc.org.br/mapasinterativos/AcidentesDeTransito/*dashboard*.html - último acesso em Abr. 11, 2019.

abordados para a presente pesquisa. Além disso, buscou-se identificar também as principais tecnologias e suas respectivas extensões utilizadas nos trabalhos analisados. Desta forma, chegou-se a lista mostrada abaixo.

- **Postgre*SQL*[30]:** é possível adicionar um complemento chamado *PostGIS*[31], o qual permite realizar análises geoespaciais. Também é possível utilizar o complemento de nome *pgRouting*[32], que permite realizar análises de rotas. A seguir são listados os trabalhos correlatos que fazem uso destas tecnologias: Simette et al. (2018), Vila et al. (2016), Kozievitch et al. (2018), Parcianello e Kozievitch (2017), Costa et al. (2017), Minetto et al. (2016), Kozievitch et al. (2017), Kozievitch et al. (2016), Diniz Junior (2017) e Barczyszyn (2015).

- **Gephi[33]:** oferece o complemento *Geolayout*[34], com o qual é possível elaborar grafos georreferenciados como, por exemplo, mapas de fluxo. Ferramenta utilizada nos trabalhos correlatos Parcianello e Kozievitch (2017) e Diniz Junior (2017).

- **Tableau[35]:** solução utilizada na elaboração do Mapa da Transparência do RS[36] e no *dashboard* de acidentes de trânsito desenvolvido pelo IPPUC[37].

- ***GMaps* API[38]:** Trabalhos correlatos que fizeram uso desta solução: Zhang et al. (2015b), Parcianello e Kozievitch (2017), Minetto et al. (2016), Kozievitch et al. (2017), Vila et al. (2016), Kozievitch et al. (2016), Diniz Junior (2017), Costa et al. (2017) e Barczyszyn (2015).

- **Mapbox[39]:** Solução utilizada nos trabalhos correlatos *City Geographics*[40] e *Manhattan Population Explorer*[41].

- **OpenStreetMap (OSM)[42]:** tecnologia utilizada nos trabalhos correlatos Simette et al. (2018), Kozievitch et al. (2018), Costa et al. (2017), Parcianello e Kozievitch

30 postgre*SQL*.org - último acesso em Set. 07, 2018.
31 *PostGIS*.net - último acesso em Set. 07, 2018.
32 *pgRouting*.org - último acesso em Set. 07, 2018.
33 gephi.org - último acesso em Set. 07, 2018.
34 gephi.org/*plugins*/#/*plugin*/geolayout-*plugin* - último acesso em Set. 07, 2018.
35 tableau.com - último acesso em Abr. 12, 2019.
36 mapa.rs.gov.br/diarias - último acesso em Abr. 11, 2019
37 ippuc.org.br/mapasinterativos/AcidentesDeTransito/*dashboard*.html – último acesso em Abr. 11, 2019.
38 developers.google.com/maps/documentation/*Javascript*/tutorial - último acesso em Set. 07, 2018.
39 mapbox.com - último acesso em Mar. 24, 2019.
40 citygeographics.org - último acesso em Mar. 24, 2019.
41 manpopex.us - último acesso em Mar. 24, 2019.
42 api.openstreetmap.org - último acesso em Mar. 10, 2019.

(2017), Minetto et al. (2016), Kozievitch et al. (2017), Vila et al. (2016), Kozievitch et al. (2016), Diniz Junior (2017) e Barczyszyn (2015).

- Leaflet[43]: biblioteca *Javascript* voltada para a criação de mapas web interativos utilizada no trabalho correlato *Manhattan Population Explorer*[44] e em Mattos et al. (2019).

- **Quantum GIS[45]:** oferece o complemento *Open Layers*[46] com o qual é possível trabalhar com mapas do *GMaps* e *OSM*. Também disponibiliza o complemento *Qgis2threejs*[47] que permite elaborar mapas georreferenciados em 3D. Solução empregada nos seguintes trabalhos correlatos: Simette et al. (2018), Kozievitch et al. (2018), Minetto et al. (2016), Kozievitch et al. (2017), Vila et al. (2016), Kozievitch et al. (2016), Diniz Junior (2017), Costa et al. (2017) e Barczyszyn (2015).

- **Rstudio[48]:** pacote *ODBC*[49] permite conectar a diferentes bancos de dados, o *SQLdf*[50] permite utilizar *Structured Query Language (SQL)* para manipular *dataset*s em geral, o *ggplot*[51] para elaborar gráficos e o *dbscan*[52] para aplicar o algoritmo de forma simples em conjuntos de dados em geral. Adotada no trabalho correlato Kozievitch et al. (2018), Parcianello e Kozievitch (2017) e Diniz Junior (2017).

2.5.6 Pesquisa de OD domiciliar

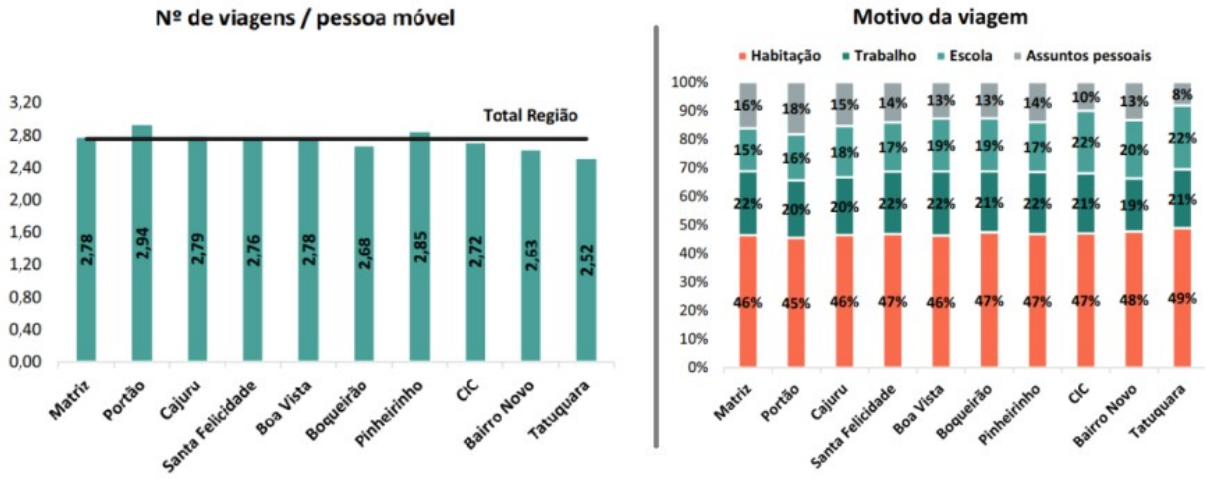
Conforme pode ser visto em IPPUC (2017), o IPPUC realizou em 2017 um estudo que visava a identificar e quantificar os principais padrões comportamentais de deslocamento dos residentes da Região Metropolitana de Curitiba (RMC) nos aspectos relacionados com mobilidade. Este inquérito domiciliar foi estruturado para dar respostas a questões relacionadas com três áreas temáticas, a saber: domicílios, população e viagens. Os dados obtidos foram analisados sob diferentes escalas geográficas: município, bairro, macro zona e

43 leafletjs.com- último acesso em Mar. 10, 2019.
44 manpopex.us - último acesso em Mar. 24, 2019.
45 qgis.org - último acesso em Set. 07, 2018.
46 *plugins*.qgis.org/*plugins*/openlayers_*plugin*/ - último acesso em Set. 07, 2018.
47 *plugins*.qgis.org/*plugins*/Qgis2threejs/ - último acesso em Set. 07, 2018.
48 rstudio.com - último acesso em Set. 07, 2018.
49 rstudio.com - último acesso em Set. 07, 2018.
50 cran.r-project.org/package=*ODBC* - último acesso em Set. 07, 2018.
51 cran.r-project.org/src/contrib/Archive/ggplot/ - último acesso em Set. 07, 2018.
52 cran.r-project.org/package=dbscan - último acesso em Set. 07, 2018.

por zona de pesquisa domiciliar. A seguir são abordadas algumas das contribuições da pesquisa realizada pelo IPPUC.

Viagens por pessoa móvel e motivação da viagem: levando-se em consideração a população de Curitiba que realiza pelo menos 01 viagem por dia, percebe-se no geral que todas as macro regiões possuem valores próximos da média regional de viagens por pessoa móvel: 2,71. Destaque para Portão com o maior valor (2,94) e para Tatuquara com o menor (2,52). Também foram analisados os motivos das viagens realizadas, os quais foram categorizados em habitação, trabalho, escola e assuntos pessoais. O alto percentual verificado na categoria habitação reflete o retorno à casa dos demais deslocamentos. Verifica-se também que os motivos de viagem apresentam distribuições equivalentes entre as macro zonas. A Figura 19 mostra os resultados obtidos para cada uma das macro regiões de Curitiba.

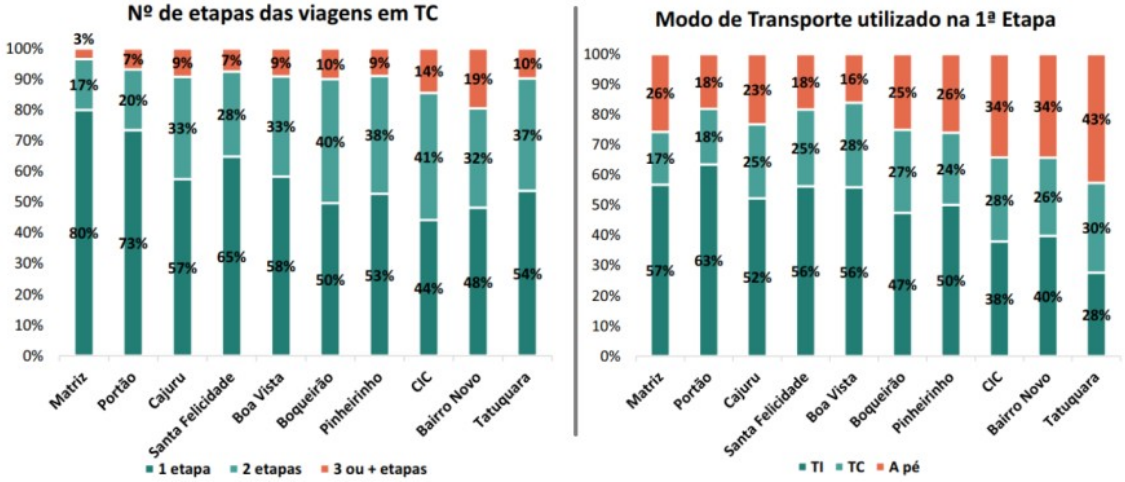
Figura 19 - Total de viagens por pessoa móvel (esq.) e motivos das viagens (dir.).



Fonte: IPPUC (2017).

As etapas das viagens: percebeu-se que a macro região Matriz possui a maior proporção de viagens de apenas 01 etapa (80%), CIC possui a maior proporção de viagens com 2 etapas (41%) e Bairro Novo o maior percentual de viagens envolvendo 3 ou mais etapas (19%). Quando perguntado qual modo de transporte era utilizado na primeira etapa das viagens, verificou-se que apenas na macro zona Tatuquara o modo dominante é o modo a pé: nas demais verificou-se um uso maior do Transporte Individual (TI) do que do Transporte Coletivo (TC). A Figura 20 mostra os resultados para cada uma das macro regiões de Curitiba.

Figura 20 - Número de etapas das viagens em TC (esq.) e modo de transporte utilizado na primeira etapa (dir.).



Fonte: IPPUC (2017).

2.6 Desafios identificados nos trabalhos correlatos

A seguir são listados os desafios identificados nos trabalhos correlatos.

- **Dados Abertos:** Kono (2016) traz uma consolidação de desafios relacionados a *Open Data*, a saber: integridade e qualidade de dados, georreferenciamento, integração de dados de diferentes provedores e gestão de metadados. Na pesquisa de Parcianello e Kozievitch (2017), por exemplo, os desafios identificados foram a não padronização da nomenclatura dos municípios extremos de viagens públicas e da falta dessa informação em alguns casos (integridade e qualidade) e a falta de localização geográfica dos *dataset*s disponibilizados pelo Portal da Transparência gaúcho (georreferenciamento). Em Beluzo (2015), os desafios identificados foram inconsistências nos dados (integridade e qualidade), problemas de incompatibilidade de *dataset*s provenientes de fontes diferentes (integração) e a falta de documentação dos *dataset*s (gestão de metadados).

- **Integração de dados heterogêneos:** o trabalho de Beluzo (2015) traz uma série de conceitos e desafios relacionados a integração de dados do ponto de vista de estrutural (dados estruturados, semi-estruturados e não-estruturados), sintático (dados idênticos nomeados de forma diferente) e semântico (dados diferentes nomeados de forma idêntica). Na pesquisa de Costa et al. (2017), por exemplo, o desafio

enfrentado foi a necessidade de integrar integrar dados contidos em arquivos de formatos *PDF*, *JSON* e *SHP* (heterogeneidade estrutural).

- **Análise espaço-temporal de dados geográficos:** o trabalho de Vila (2016) menciona o desafio relacionado a pouca ou inexistente disponibilização histórica de *dataset*s por parte dos provedores de dados georreferenciados. A pesquisa de Kozievitch et al. (2017) e o projeto *Manhattan Population Explorer*[53] são exemplos de trabalhos evidenciam a relevância da disponibilização histórica de dados em SIG.

- **Sistemas de Informação Geográfica baseados em *crowdsourcing*:** soluções baseadas em *crowdsourcing* podem apresentar uma série de desafios. [Harris, 2018] relata que a abertura demandada pela adoção de *crowdsourcing* faz com que a solução se torne vulnerável à sabotagem por parte de atores desonestos ou competitivos. Neste sentido, Kono (2016) acrescenta que aplicativos que dependem de dados *crowdsourced* apresentam inerentemente problemas de segurança e de qualidade de dados.

- **Problemas envolvendo análise OD:** Um desafio relacionado a problemas do tipo OD é quando não se dispõe dados de origem e/ou de destino explicitamente definidos. Nestes casos, costumam-se utilizar diferentes técnicas para deduzir a origem e o destino dos deslocamentos sob análise. Isto foi feito na pesquisa de Diniz Junior (2017), que inclusive buscou mitigar a imprecisão da metodologia adotada através da clusterização dos dados obtidos.

- **Visualização de grandes volumes de dados geográficos:** Têm sido frequente a utilização de técnicas de clusterização de dados, processo através do qual é possível abstrair detalhes do *dataset*, diminuir o volume de dados sob análise e simplificar a visualização dos dados. Ainda, clusterização aplicada a SIG propicia a realização de análises sob diferentes níveis de detalhamento, cujas consultas tendem a ter um menor tempo de resposta do que se realizadas contra o *dataset* em sua granularidade natural. Como exemplos de trabalhos que adotaram abordagens neste sentido podemos citar o de Diniz Junior (2017), Lu et al. (2015), Zhang et al. (2015b), Vila (2016) e o projeto Taxivis[54].

53 manpopex.us - último acesso em Mar. 24, 2019.
54 github.com/ViDA-NYU/TaxiVis - último acesso em Mar. 24, 2019.

3 PROTÓTIPO PARA VISUALIZAÇÃO DE OD

Neste capítulo é apresentado o processo de desenvolvimento do protótipo proposto, desde os trabalhos de levantamento de requisitos, a construção da base de dados, o *tuning* das consultas *SQL* e da base de dados, a obtenção das regiões de interesse, a arquitetura da aplicação, as tecnologias utilizadas, a interface de usuário e as lições aprendidas no decorrer deste processo.

3.1 O levantamento de requisitos

Baseado nos trabalhos de Ferreira et al. (2011), Ferreira et al. (2013), Lu et al. (2015), Zhang et al. (2015b), Vila (2016) e Diniz Junior (2017), que envolvem o desenvolvimento de solução para visualização de dados, decidiu-se desenvolver uma ferramenta que permitisse visualizar os embarques e desembarques dos usuários do STP de Curitiba. O objetivo pretendido através da ferramenta seria apoiar o processo de compreensão de aspectos relacionados ao perfil do usuário do transporte público, a variação da demanda do STP ao longo de um determinado período, os locais onde ocorrem a maior quantidade de embarques e desembarques, compreender a dinâmica de ocupação do espaço urbano, dentre outros.

Para verificar a relevância desta proposta, elaborou-se um questionário baseado em abordagens verificadas no Relatório 5 - Pesquisa de Origem-Destino domiciliar de Curitiba IPPUC (2017), no qual análises de OD são realizadas por período, sob o ponto de vista de bairros e regionais de embarque e desembarque, e inclusive por sexo e faixa etária dos usuários. O questionário completo resultante pode ser visto no Apêndice A, que foi aplicado a um grupo de 7 pesquisadores da área de mobilidade urbana durante os dias 17 a 24 de abril de 2019. O referido instrumento trazia questões como: a) se o respondente já havia tido algum contato com estudos relacionados ao uso do STP de Curitiba, b) se seria interessante desenvolver uma solução que permitisse visualizar, quantificar e explorar dados do STP de Curitiba, c) quais filtros de pesquisa a solução poderia oferecer e d) quais seriam as possibilidades de uso da solução. A partir da análise das respostas obtidas, verificou-se que todos os respondentes já haviam tido contato com estudos envolvendo o uso de STP de Curitiba. Os respondentes apontaram também que que seria interessante dispor de uma

solução que permitisse visualizar, quantificar e explorar dados relacionados ao uso do transporte público. A Figura 21 traz uma síntese das respostas obtidas quando perguntado quais opções de filtro de pesquisa uma solução de visualização de dados de Origem-Destino deveria oferecer: os números mostrados ao lados das barras horizontais representam a quantidade de votos que cada uma das possibilidades de filtros recebeu.

Figura 21 - Filtros desejados em uma solução de visualização de dados de Origem-Destino.

Fonte: do próprio autor.

3.2 A base de dados da aplicação

Os dados usados neste trabalho foram obtidos a partir da Prefeitura Municipal de Curitiba (PMC)[55] e do Instituto de Pesquisa e Planejamento Urbano de Curitiba (IPPUC)[56], totalizando cerca de 17GB de dados em formato *SHP* e *CSV*. A lista a seguir traz uma breve descrição dos conjuntos de dados iniciais utilizados no protótipo: todos são dados abertos, exceto os 2 últimos que foram fornecidos pela Urbanização de Curitiba (URBS). Na sequência, a Tabela 2 apresenta as principais características observadas em tais dados.

- **Terminais rodoviários:** conjunto georreferenciado de dados em formato *SHP* com 24 registros referentes aos terminais rodoferroviários do STP de Curitiba.
- **Pontos de ônibus:** conjunto georreferenciado de dados em formato *SHP* com 16590 registros referentes a localização dos pontos de ônibus do STP de Curitiba.
- **Contornos dos bairros:** conjunto georreferenciado de dados em formato *SHP* com 75 registros referentes aos contornos dos bairros de Curitiba.

55 www.curitiba.pr.gov.br/dadosabertos/ - último acesso em Set. 07, 2018.
56 ippuc.org.br/geodownloads/geo.htm - último acesso em Set. 07, 2018.

- **Contornos das regionais:** conjunto georreferenciado de dados em formato *SHP* com 10 registros referentes aos contornos das regionais de Curitiba.

- **Registros dos cartões:** conjunto de dados não georreferenciados que continha os registros de utilização dos cartões de transporte dos usuários do STP referentes a 31 dias de operação do STP (do dia 30/09/2017 ao dia 30/10/2017). Os dados foram disponibilizados em 31 arquivos de texto com extensão *TXT* e conteúdo em formato *CSV*. Cada arquivo continha aproximadamente 70*MB* de dados para cada dia de operação (cerca de 280.000 registros por arquivo, em média).

- **Registros das posições dos ônibus:** conjunto de dados georreferenciados que continha a localização dos ônibus do STP em operação durante 31 dias (do dia 30/09/2017 ao dia 30/10/2017), dados estes coletados a cada 5 minutos, em média. Os dados foram disponibilizados em 31 arquivos de texto com extensão *TXT* e conteúdo em formato *CSV*. Cada arquivo continha aproximadamente 500*MB* de dados para cada dia de operação (cerca de 3,5 milhões de registros por arquivo, em média).

Tabela 2 - Algumas informações sobre os dados iniciais utilizados nesta pesquisa.

Descrição	Term.	Pontos ônibus	Bairros	Macror.	Cartões usuários	Posição ônibus
Quantidade total de arquivos	1	1	2	1	31	31
Formato do arquivo	*SHP*	*SHP*	*SHP*	*SHP*	*TXT*	*TXT*
Quantidade total de registros	24	16.590	75	10	8.710.082	114.602.481
Georreferenciado?	Sim	Sim	Sim	Sim	Não	Sim
Quantidade de veículos identificados	-	-	-	-	1.481	1.575
Quantidade de linhas de ônibus	-	-	-	-	286	355

Uma vez inseridos no Postgre*SQL*, iniciou-se uma série de procedimentos e cruzamentos de dados para viabilizar as tabelas necessárias para o desenvolvimento do protótipo. A lista seguinte traz uma breve descrição de cada uma destas tabelas:

- **ROI:** acrônimo de *ROI - Regions Of Interest* (Regiões de Interesse em inglês), nesta tabela estão reunidos os contornos dos bairros e das regionais de Curitiba.

Também contém os contornos dos *clusters*, os quais são abordados com maiores detalhes na Seção 3.3.

- **Terminais e Pontos de Ônibus:** tabela que reúne os dados referentes a localização dos pontos de ônibus e dos terminais de ônibus.

- **Movimentação:** tabela resultante do cruzamento dos registros de utilização dos cartões de usuários com os de posição dos veículos do STP de Curitiba. Cada tupla refere-se a uma viagem realizada por um determinado usuário do STP: contém as coordenadas de localização dos extremos do deslocamento, data-hora destes instantes, código da linha e do veículo relacionado, data de nascimento, idade, faixa etária e sexo do usuário. Nos apêndices C e D estão disponíveis maiores detalhes sobre o processo de dedução dos pontos de embarque e desembarque, respectivamente, os quais seguem a abordagem adotada por Diniz Junior (2017).

A Figura 22 mostra a estrutura dos dados iniciais e da base resultante, cujo dicionário de dados pode ser verificado no Apêndice E.

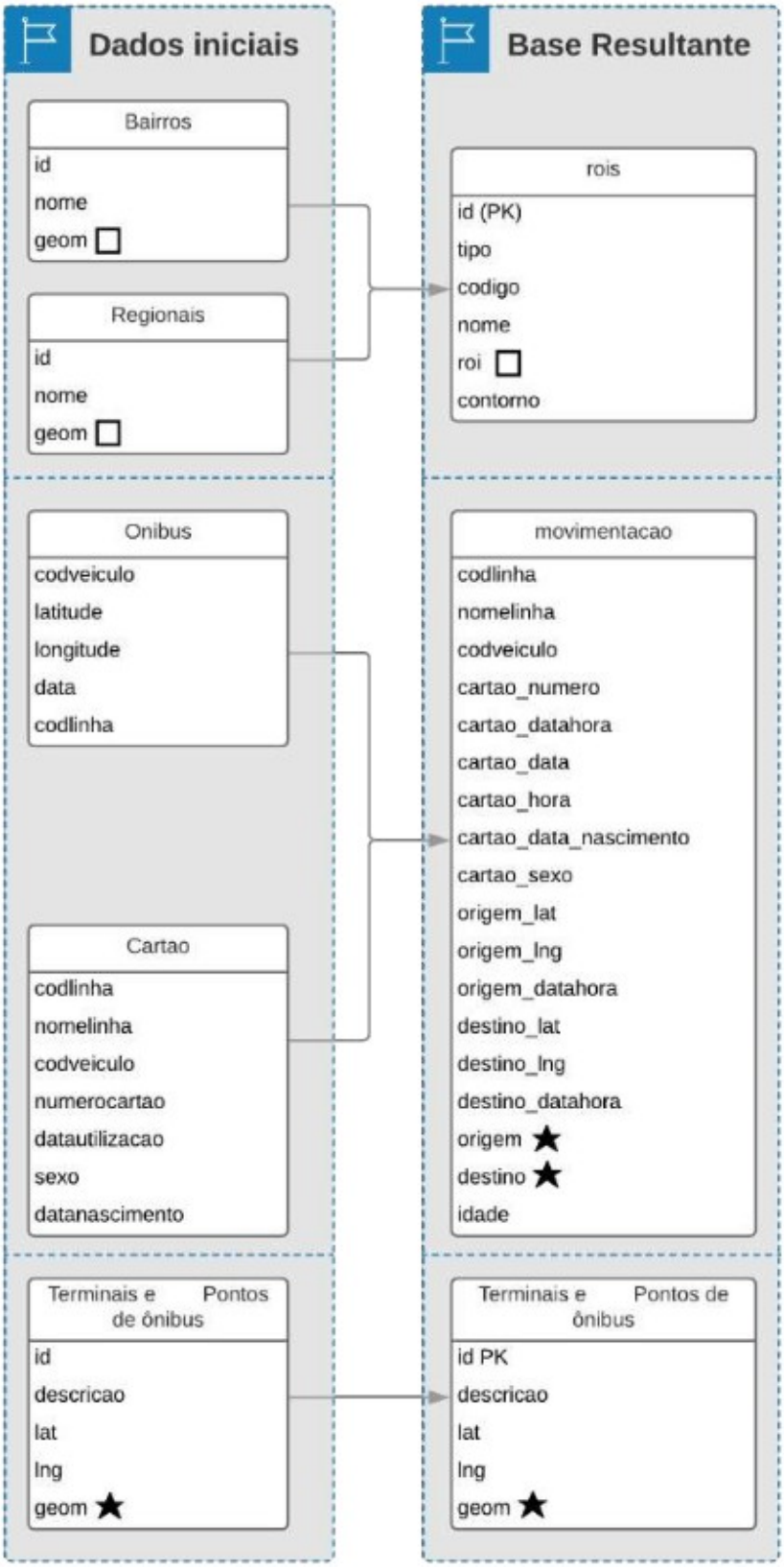
3.3 As regiões de interesse oferecidas

No protótipo, buscou-se permitir analisar os deslocamentos dos usuários do STP de Curitiba sob o ponto de vista de divisões geográficas. Para tanto, baseou-se nos contornos dos bairros e das macrorregiões da cidade. A Figura 23 mostra estas possibilidades.

Além disso, motivados pelos trabalhos de Osama et al. (2015), Silva et al. (2016), Dias et al. (2016) e Vila et al. (2016), optou-se por utilizar o algoritmo de clusterização *k-means* para clusterizar os dados a partir das informações de latitude e de longitude dos extremos dos deslocamentos (ou seja, o ponto de embarque e o de desembarque de uma viagem de um usuário do STP) e, com isso, ofertar ao usuário não apenas regiões de interesse de origem geográfica (bairros ou regionais), mas também outras de origem matemática. Desta forma, cada *cluster* obtido através do algoritmo *k-means* torna-se uma Região de Interesse (*ROI*, sigla em inglês) a partir das quais podem ser realizados novos estudos do STP, abordagem esta adotada nos trabalhos de Lu et al. (2015) e Zhang et al. (2015b), por exemplo.

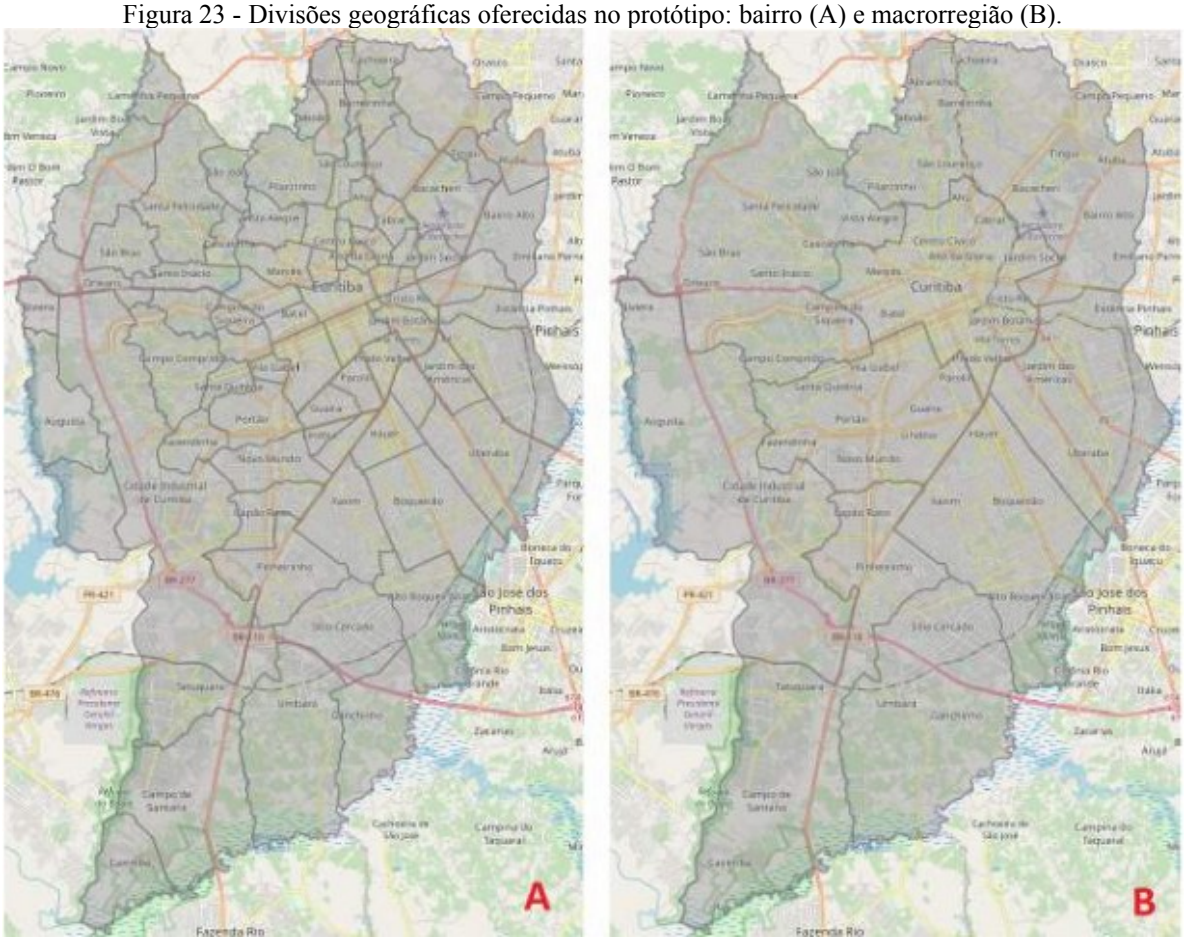
Para a obtenção dos *clusters* foi utilizado o *software RStudio* v. 1.1.463 combinado com os seguintes recursos: 1) biblioteca "*RpostgreSQL*" que permitiu conectar o *RStudio* ao banco *PostgreSQL*, 2) biblioteca "*SQLdf*" que permitiu disparar sentenças *SQL* a partir do *RStudio* ao banco *PostgreSQL* e 3) função "*kmeans*" já nativa no *RStudio*, para a qual se

Figura 22 - Estrutura dos dados iniciais e da base resultante utilizada pelo protótipo. A semântica geográfica dos campos espaciais envolvidos é informada conforme o padrão *Object Modeling Technique for Geographic Applications (OMT-G)*.



Fonte: do próprio autor.

Figura 23 - Divisões geográficas oferecidas no protótipo: bairro (A) e macrorregião (B).



Fonte: do próprio autor.

informa a quantidade de *clusters* desejada e a lista de dados a serem clusterizados e cujo retorno é uma lista dos *clusters* nos quais cada elemento da lista original foi alocado. Desta forma, utilizou-se uma janela de 07 dias consecutivos de dados de embarque e desembarque de usuários e o *k-means* foi executado para classificar tais dados em 5, 10, 20, 30, 40, 50, 100, 250, 500 e 1000 diferentes *clusters*. O *script* escrito em linguagem *R* e utilizado para obter tais conjuntos de *clusters* pode ser visto no Apêndice H.

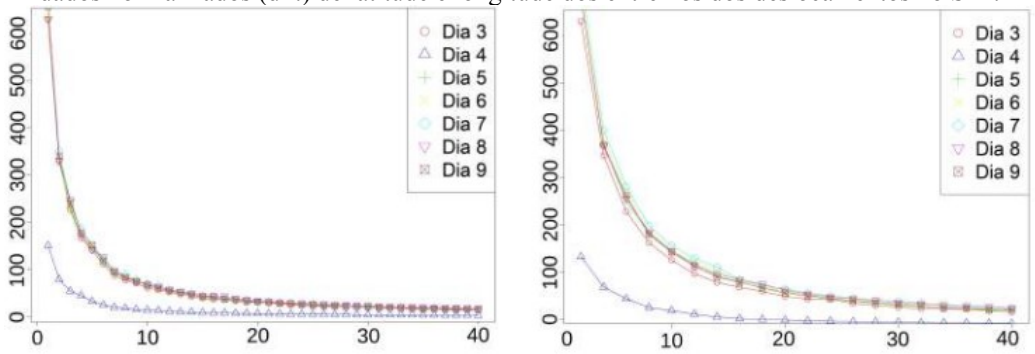
De posse dos *clusters* obtidos via *k-means*, o passo seguinte foi obter os contornos que delimitavam cada um destes grupos. Para tanto, foi utilizada a função *ST_convexhull* do *PostGIS*, cujo funcionamento equivale a envolver com um elástico um conjunto de pontos georreferenciados. Assim, ao conjunto dos contornos dos bairros e das regionais de Curitiba, foram adicionados os contornos dos *clusters* dos deslocamentos de usuários do STP daquela cidade mostrados na Figura 24. Observa-se na Figura 24 que quanto maior o valor de *k* (equivalente a quantidade de *clusters*), maiores são as áreas territoriais de Curitiba não cobertas pelos *clusters*.

Figura 24 - Diferentes *ROI* oferecidas pela aplicação.

Fonte: do próprio autor.

Cabe ressaltar que foram realizadas tentativas para estimar a quantidade ótima de *clusters* a serem produzidos: para tanto, foi utilizado o *Elbow Method*. Na Figura 25 são exibidos os resultados obtidos: o gráfico da esquerda mostra os resultados obtidos para cada dia da semana quando analisados os dados de latitude e de longitude na sua forma original; o da direita foi obtido quando analisados os dados normalizados (processo de normalização realizado através da utilização da função *rnorm*, já nativa no *Rstudio* v. 1.1.463). Em ambos os casos, buscou-se variar o valor de *k* desde 2 até 40 (vide *script* no Apêndice I). Tais resultados não foram suficientes para indicar um número ótimo de *clusters*, uma vez que não ficou evidenciada a formação de um cotovelo nas curvas dos gráficos.

Figura 25 - Gráfico obtido via *Elbow Method* para cada dia da semana a partir dos dados brutos (esq.) e dados normalizados (dir.) de latitude e longitude dos extremos dos deslocamentos no STP.



Fonte: do próprio autor.

3.4 Desempenho das consultas

Nossa solução de visualização de dados de Origem-Destino foi desenvolvida levando-se em consideração as respostas coletadas através da aplicação do questionário apresentado na Seção 3.1. Nesta direção, buscou-se adicionar ao protótipo a possibilidade de analisar tais dados via aplicação dos seguintes filtros de dados: por data e hora de embarque, por sexo e idade do usuário, por região de embarque e de desembarque.

Inicialmente buscou-se averiguar o tempo de resposta de consultas que poderiam ser disparadas a partir do protótipo contra a base de dados. Para tanto, foram elaboradas 12 sentenças *SQL*, as quais representavam todas as possíveis combinações de filtros que seriam oferecidos pelo protótipo. Tais consultas foram disparadas contra o banco e os respectivos tempos de execução foram contabilizados. As consultas utilizadas podem ser visualizadas no Apêndice F.

Na sequência, buscou-se otimizar o tempo de execução das referidas consultas através da criação de índices do tipo *B-Tree*. A Tabela 3 traz os tempos de execução de cada uma das consultas disparadas antes e após a criação de tais índices. Verificou-se também que os índices providenciados resultaram em uma redução média de 57% do tempo de execução das consultas.

3.5 Arquitetura

O protótipo foi concebido com o intuito de permitir que usuários pudessem facilmente realizar análises de OD sem que precisassem fazer uso de qualquer tipo de linguagem de programação e/ou de manipulação de dados. Também optou-se por desenvolver uma solução de maneira que não se fizesse necessário, por parte do usuário final, nenhum tipo de instalação de quaisquer *plugins* para utilização do protótipo. Ainda, buscou-se utilizar apenas tecnologias *open-source* como forma de evitar gastos com aquisição e/ou contratação de *softwares*. Foram utilizadas as seguintes tecnologias:

- Banco de dados *PostgreSQL* v. 9.5 x64[57] e extensão espacial *PostGIS* v. 2.1.1[58];

57 www.postgre*SQL*.org/download/linux/ubuntu/ - último acesso em Mar. 10, 2019.
58 *PostGIS*.net/2013/11/08/*PostGIS*-2.1.1/ - último acesso em Mar. 10, 2019.

Tabela 3 - Tempos de execução de sentenças *SQL* antes e após a criação de índices.

Consulta disparada	Tipo de *ROI* utilizado	Qtd de origens selecionadas	Qtd de destinos selecionados	Qtd de dias	Intervalo de horários	Sexo selecionado	Faixa etária selecionada	Qtd de tuplas retornadas	Tempo (s)		Redução do tempo de resposta
									Sem índices	Com índices	
Consulta A	Bairro	1	1	7	00:00 a 23:59	-	-	994	19	8	57,89%
Consulta B	Bairro	1	1	7	00:00 a 23:59	F	-	536	30	4	86,67%
Consulta C	Bairro	4	4	7	00:00 a 23:59	-	-	1890	44	11	75,00%
Consulta D	Bairro	4	4	7	00:00 a 23:59	F	-	1062	28	9	67,86%
Consulta E	Regional	1	1	15	00:00 a 20:00	-	-	7892	49	29	40,82%
Consulta F	Regional	1	1	15	00:00 a 20:00	F	-	4713	60	28	53,33%
Consulta G	Regional	1	1	15	00:00 a 20:00	-	18 a 65	6190	36	27	25,00%
Consulta H	Regional	1	1	15	00:00 a 20:00	F	18 a 65	3662	60	23	61,67%
Consulta I	Regional	2	2	30	00:00 a 23:59	-	-	32773	1020	480	52,94%
Consulta J	Regional	2	2	30	00:00 a 23:59	F	-	19152	600	240	60,00%
Consulta K	Regional	2	2	30	00:00 a 23:59	-	18 a 65	25259	780	360	53,85%
Consulta L	Regional	2	2	30	00:00 a 23:59	F	18 a 65	14680	420	180	57,14%

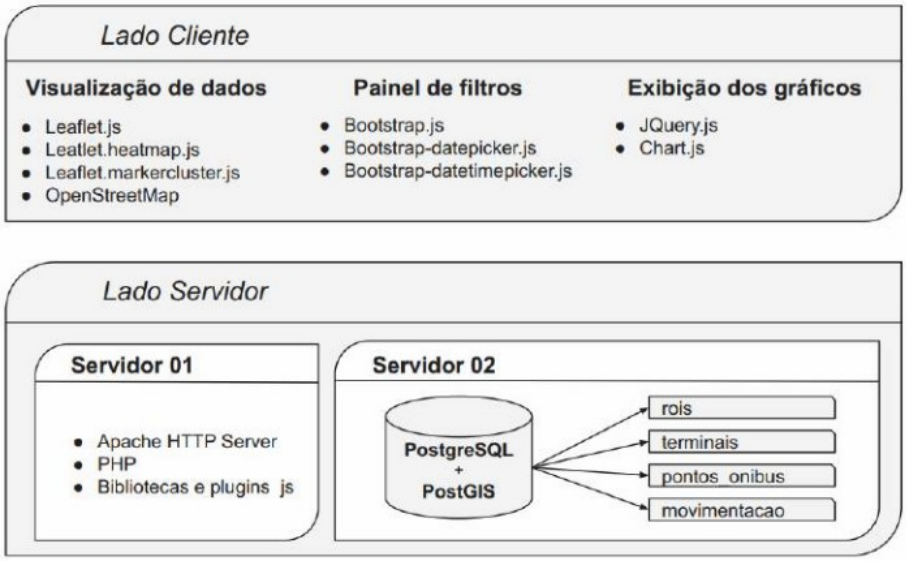
- Servidor *web Apache Server* v. 2.4.37[59] e *PHP* v. 7.2.12[60];
- Base de mapas do *Open Street Map*[61] combinada com a biblioteca *Leatlet* v. 0.7.7[62] e *plugins Markercluste*r v. 0.5.0[63] e *Heatmap* v. 0.7.7[64]
- Bibliotecas *JQuery* v. 3.3.1[65], *Bootstrap* v. 3.3.7[66], *Datepicker* v. 1.6.4[67], *Datetimepicker* v. 4.17.47[68] e *Chart.js* v. 2.7.3[69].

Todas essas tecnologias foram hospedadas em 02 diferentes servidores, cujas principais características são descritas a seguir:

- Servidor 01 (de aplicação): *Intel Core i*7 3632*QM* 4-*core* 8-*threads* x64 2.2*GHz*, 8*GB RAM*, *Windows* 7 64 *bits*;
- Servidor 02 (de dados): *AMD EPYC* 7401 24-*Core* 48-*threads* x64 2.8*GB*, *Debian* 9.9 64 *bits*.

A Figura 26 mostra um diagrama de como tais tecnologias foram combinadas para o desenvolvimento do protótipo. O código-fonte da aplicação está disponível no *Github*[70]. O apêndice G traz uma breve descrição dos principais arquivos da aplicação.

Figura 26 - Tecnologias utilizadas na solução desenvolvida.



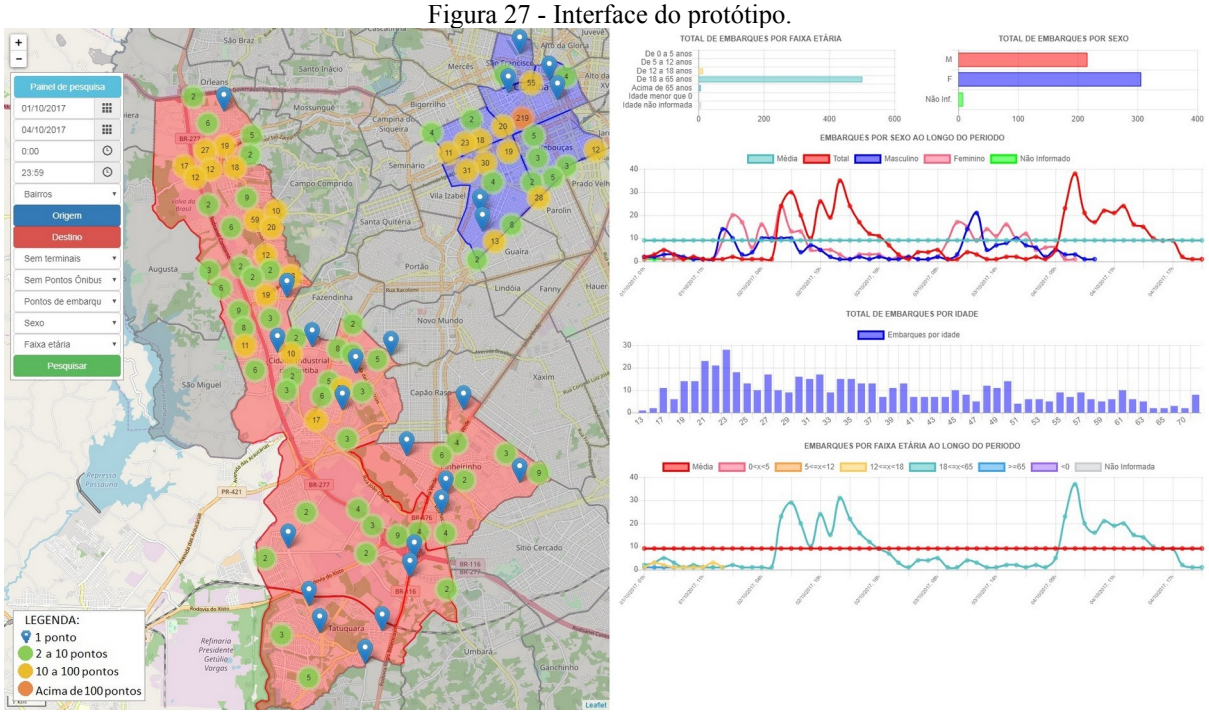
Fonte: do próprio autor.

59 archive.apache.org/dist/httpd/ - último acesso em Mar. 10, 2019.
60 php.net/releases/7_2_12.php - último acesso em Mar. 10, 2019.
61 www.openstreetmap.org - último acesso em Mar. 10, 2019.
62 leafletjs.com/download.html - último acesso em Mar. 10, 2019.
63 github.com/Leaflet/Leaflet.*Markercluster* - último acesso em Mar. 10, 2019.
64 github.com/Leaflet/Leaflet.heat - último acesso em Mar. 10, 2019.
65 jquery.com/ - último acesso em Mar. 10, 2019.
66 getbootstrap.com - último acesso em Mar. 10, 2019.
67 github.com/eternicode/bootstrap-datepicker - último acesso em Mar. 10, 2019.
68 github.com/Eonasdan/bootstrap-datetimepicker - último acesso em Mar. 10, 2019.
69 www.chartjs.org/ - último acesso em Mar. 10, 2019.
70 https://github.com/yussefparcianello/OrigemDestinoStpCuritiba - último acesso em Out. 20, 2019.

3.5.1 A interface da aplicação

No protótipo, toda interação ocorre em uma única tela. Conforme pode ser visto na Figura 27, a interface da ferramenta é dividida verticalmente ao meio: à esquerda é disponibilizado um mapa interativo com um painel de pesquisa; à direita são disponibilizados os resultados da consulta realizada. É através do painel de pesquisa que são oferecidos filtros para consulta, estabelecidos com base na análise das respostas do questionário mencionado previamente.

Figura 27 - Interface do protótipo.



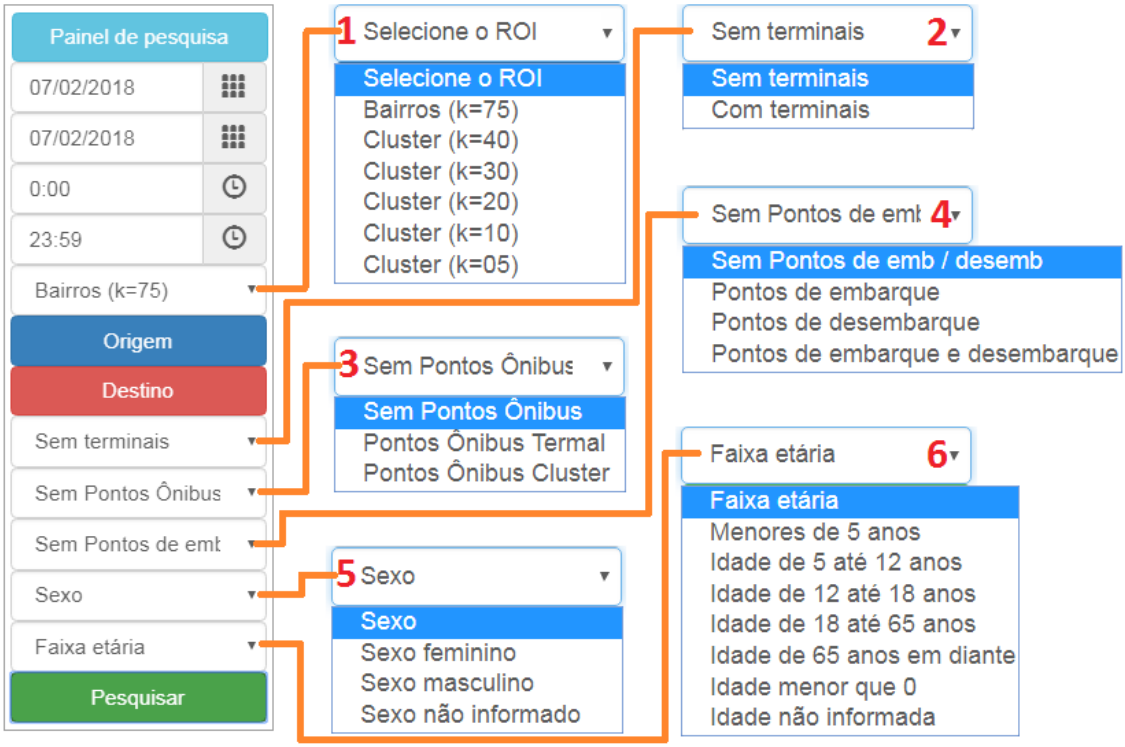
Fonte: do próprio autor.

A Figura 28 dá ênfase aos filtros do painel de pesquisa. Os quatro primeiros campos de filtro permitem definir a data inicial, a data final, o horário inicial e o horário final do período a ser analisado, respectivamente. O campo 1 oferece uma caixa de seleção através da qual o usuário informa como deseja realizar a pesquisa: a partir de divisões geográficas (bairros ou regionais) ou por algum dos conjuntos de *clusters* de origem matemática.

Para selecionar uma origem, o usuário aciona o botão "Origem" (em azul) e então seleciona as regiões desejadas, os quais serão coloridas com a cor azul. De forma análoga, para a seleção de destino, aciona-se o botão "Destino" e seleciona-se as regiões desejadas, as quais terão suas cores alteradas para vermelho. É possível realizar análises do tipo intra-

regiões bastando, para isso, selecionar a mesma área como origem e destino. O mapa apresenta os principais pontos de origem e destino dos deslocamentos que satisfazem os critérios de busca. Os demais campos identificados com os números 2, 3 e 4 (terminais, pontos ônibus e pontos de embarque/ desembarque, respectivamente) permitem adicionar camadas ao mapa de maneira que sirvam como subsídios para que o usuário possa melhor decidir quais regiões do mapa explorar.

Figura 28 - Filtros do Painel de Pesquisa do protótipo.



Fonte: do próprio autor.

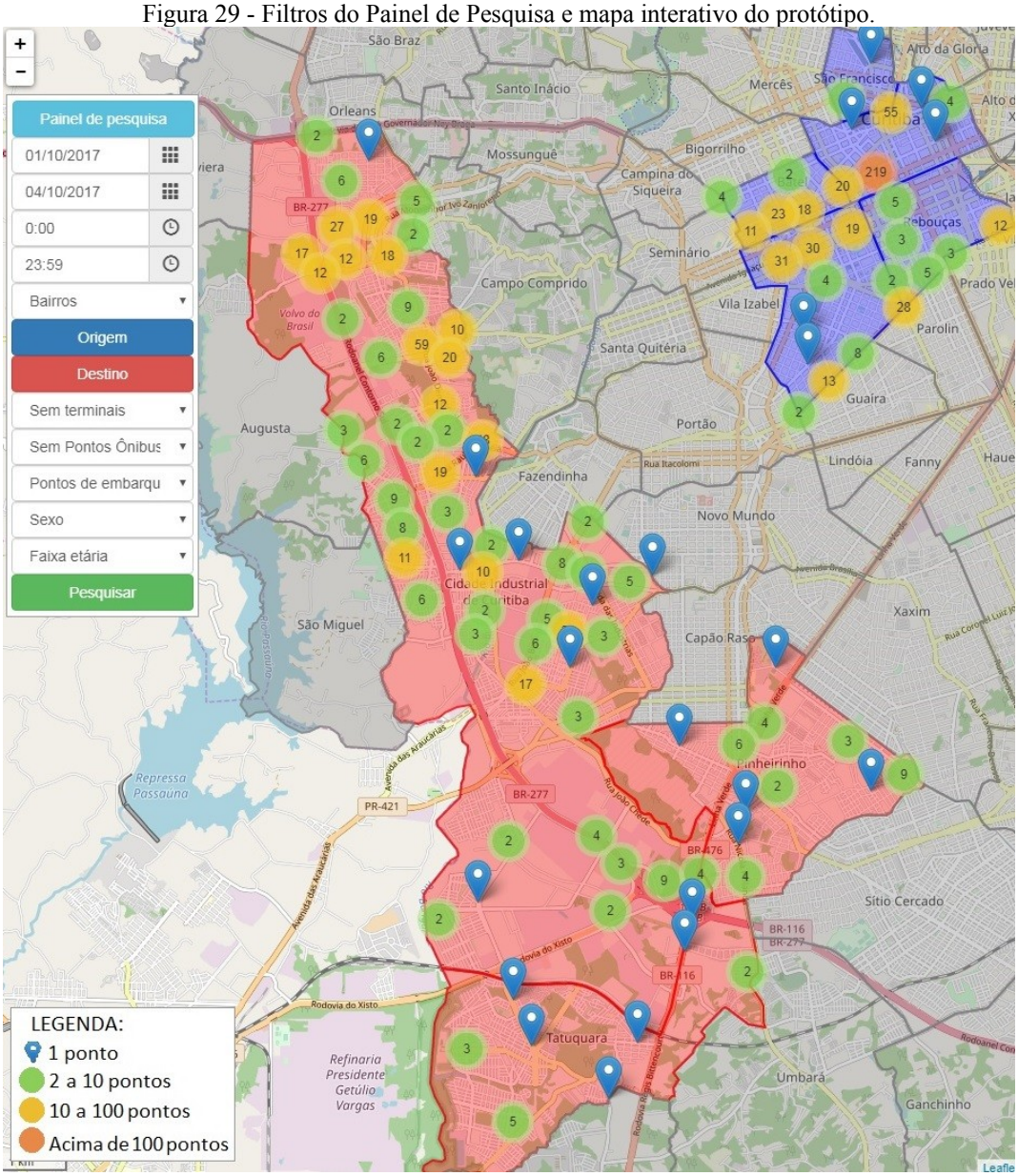
A ferramenta permite também analisar os deslocamentos com base no sexo dos usuários: vide a opção 5 da Figura 28. É possível também realizar pesquisas com base em faixas etárias, as quais foram estabelecidas com base em diferentes situações. As faixas etárias "Menores de 5 anos" e "Idade de 65 anos em diante" foram estabelecidas em razão da isenção tarifária no transporte público de Curitiba. Já a faixa etária "Idade de 5 até 12 anos" visa a contemplar apenas crianças, de acordo com o artigo 2 do Estatuto da Criança e do Adolescente (ECA)[71] cuja idade não faz jus a isenção tarifária no STP de Curitiba. A faixa etária "Idade de 12 até 18 anos" foi estabelecida também com base no ECA (artigo 2) para contemplar apenas o público adolescente. Já a faixa etária "Idade de 18 até 65 anos" visa a contemplar os demais usuários do transporte público. Finalmente, as faixas etárias "Idade

71 http://www.planalto.gov.br/ccivil_03/leis/l8069.htm – último acesso em Fev. 15, 2020.

menor que 0" e "Idade não informada" permitem explorar os casos de usuários do STP cuja informação referente a idade encontra-se incorreta ou ausente na base de dados do IPPUC.

Na Figura 29, por exemplo, é possível visualizar de forma georreferenciada o resultado de uma consulta envolvendo os deslocamentos de usuários do STP que embarcaram nos bairros marcados em azul e que desembarcaram nas regiões marcados em vermelho no intervalos de data e hora especificados.

Figura 29 - Filtros do Painel de Pesquisa e mapa interativo do protótipo.



Fonte: do próprio autor.

Os resultados obtidos através desta consulta são mostrados na Figura 30. O gráfico mostrado na Figura 30-A exibe o total de embarques realizados por faixa etária: no caso sob estudo, percebe-se que a grande maioria dos usuários está na faixa dos 18 aos 65 anos. O gráfico mostrado na Figura 30-B mostra o total de embarques por sexo: no caso em questão,

nota-se que os homens se deslocam mais que as mulheres entre os extremos envolvidos. O gráfico mostrado na Figura 30-C permite visualizar o volume de embarques por sexo realizados ao longo de todo o período sob análise: percebe-se uma série de máximos locais referentes aos momentos de pico de uso do STP ao longo dos dias analisados. O gráfico mostrado na Figura 30-D, é um histograma de frequência do volume de embarques por idade: percebe-se uma série de embarques realizados por pessoas com idade menor que zero, indicando uma inconsistência nos dados disponibilizados. O gráfico mostrado na Figura 30-E, o de linhas, permite visualizar o volume de embarques por faixa etária realizados ao longo de todo o período. Um vídeo demonstrando o funcionamento da ferramenta pode ser visto no *YouTube*[72].

Figura 30 - Resultados da pesquisa mostrada na Figura 29.

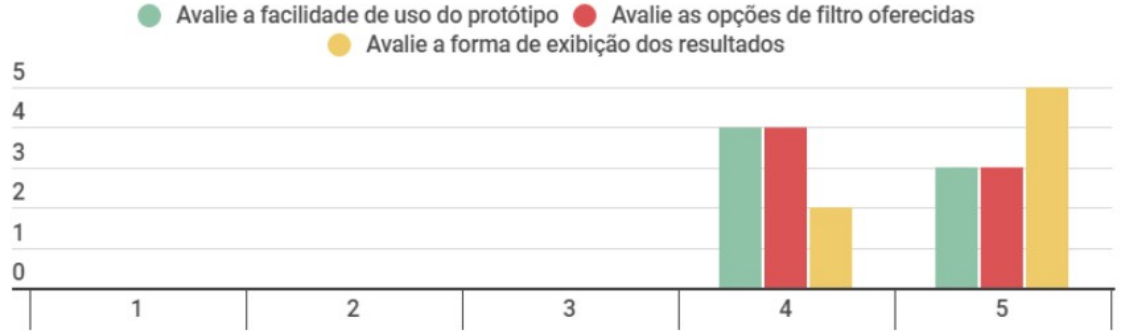
Fonte: do próprio autor.

72 https://www.youtube.com/watch?v=KOzFHRc7lXA - último acesso em Out. 20, 2019.

3.5.2 Teste de usabilidade

Percepções de usuários sobre o uso da nossa solução foram coletadas através da realização de um teste de usabilidade. Para tanto, elaborou-se um roteiro composto por 03 atividades que deveriam ser realizadas pelos voluntários no nosso protótipo. O roteiro era composto por uma tarefa fácil, uma tarefa de média complexidade e uma terceira tarefa considerada a complexa das três. O roteiro completo encontra-se no Apêndice K. Participaram do teste 07 voluntários com diferentes perfis profissionais: haviam formandos do curso de Sistemas de Informação da UTFPR, uma professora doutoranda da área de Sociologia, um professor mestre em Matemática e uma professora mestra em Geografia. Após a realização das atividades contidas no roteiro, os voluntários foram convidados a preencher o formulário de avaliação disponível no Apêndice J. O instrumento era composto por 05 questões, sendo 04 delas fechadas e 01 aberta do tipo discursiva. Todos os voluntários foram capazes de realizar na íntegra todas as tarefas do roteiro e avaliaram de forma positiva a nossa solução. A Figura 31 mostra algumas das percepções dos voluntários quanto a experiência de utilização do nosso protótipo.

Figura 31 - Percepções dos voluntários que participaram do teste de uso do protótipo.



Fonte: do próprio autor.

3.5.3 Lições Aprendidas

A realização deste trabalho evidenciou o fato de que o desenvolvimento de uma solução voltada para visualização de dados de OD envolve uma série de desafios, quais sejam:

- Dos relacionados a construção da interface, notou-se a importância de se dispor de respostas para questões como: quais filtros de pesquisa ofertar ao usuário? como implementar tais filtros? como disponibilizar estes filtros de forma que o usuário seja

capaz de utilizá-los de forma intuitiva? de qual forma os dados resultantes dessas buscas serão apresentados ao usuário?

- Para responder as perguntas, é crucial definir o público-alvo da solução. Para tanto, é pertinente buscar respostas para perguntas como: quem será o usuário da aplicação? Qual o nível de experiência do usuário no uso das tecnologias envolvidas? o usuário é capaz de compreender os resultados na forma e densidade em que serão apresentados?

- Embora a realização da análise exploratória seja fundamental para caracterizar os dados sob estudo, vê-se oportuno desenvolver soluções de OD partindo-se não só dos dados sobre os dados, mas principalmente do ponto de vista do usuário. Nesta direção, é oportuno refletir sobre o quão natural, intuitivo e fluído será ao usuário utilizar a ferramenta e o quão relevante e agregador serão para o usuário as informações fornecidas pela aplicação;

- O tempo de resposta de uma ferramenta deste tipo é também outro desafio face ao volume de dados, arquitetura, tecnologias e consultas envolvidas. Na nossa solução, foram utilizadas diferentes estratégias para melhorar o tempo de resposta da aplicação, tais como a definição dos tipos de dados das tabelas com olhos voltados para o ganho de desempenho, otimização de consultas e criação de diferentes índices, particionamento de tabelas e a utilização de Regiões de Interesse como forma de restringir os dados envolvidos em consultas a somente aqueles inseridos em tais regiões.

- No que diz respeito às tecnologias envolvidas na concepção de uma solução deste tipo, o desafio centra-se em integrar diferentes tecnologias, principalmente quando de trata de *plugins Open Source*. Por questões de incompatibilidade de recursos, muitas vezes quando se deseja adicionar uma nova funcionalidade ao sistema, por exemplo, é necessário redesenhar grande parte da aplicação para substituir *plugins* conflitantes por outros que não causem tal problema. No nosso protótipo, alguns *plugins* adotados do *Leaflet* se mostraram incompatíveis com outros que pretendíamos também adicionar no protótipo.

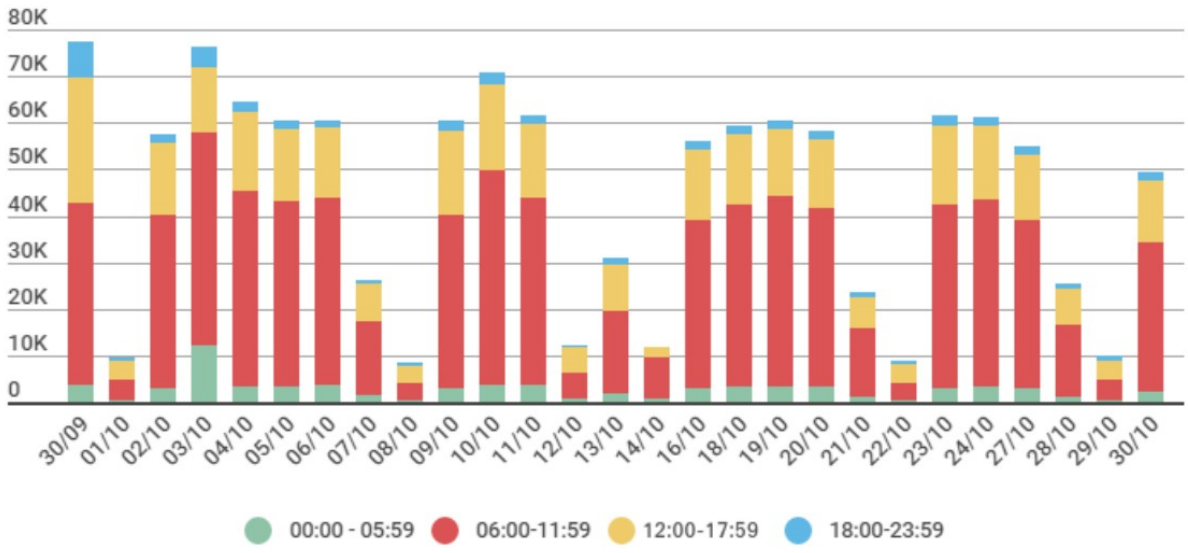
4 ESTUDO DE CASO

Este capítulo aborda os seguintes assuntos: análise exploratória dos dados do STP de Curitiba e estudo da dinâmica de ocupação do espaço urbano baseado na movimentação de usuários do STP de Curitiba.

4.1 Demanda do STP de Curitiba

Inicialmente buscou-se compreender como a demanda do STP de Curitiba varia ao longo do período analisado sob o ponto de vista de turnos em um dia. A Figura 32 apresenta o número de embarques diários: observa-se uma maior demanda nos dias úteis, e uma diminuição nos finais de semana. Independente do dia, a demanda é maior durante os turnos manhã, seguido da tarde, madrugada e, com a menor demanda, o turno da noite.

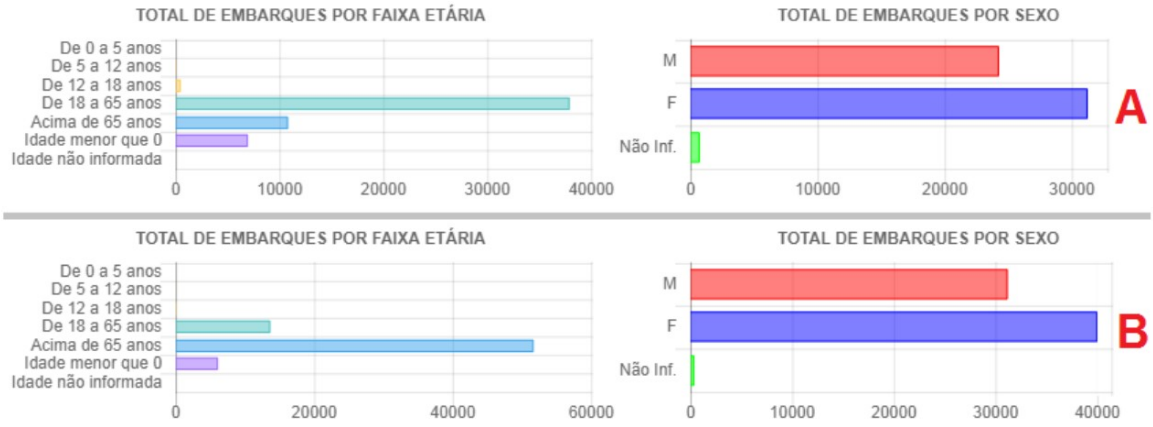
Figura 32 - Número de embarques registrados desde 30/09 até 30/10/2017.



Fonte: do próprio autor.

A partir de tais informações, utilizar o nosso protótipo para investigar qual seria o perfil dos usuários que utilizam o transporte público no período da noite e madrugada. Os resultados obtidos indicaram que são as mulheres quem utilizam mais o transporte público em tais períodos, com a diferença que no turno da noite o predomínio é de mulheres adultas (vide Figura 33-A) e na madrugada de mulheres idosas (vide Figura 33-B).

Figura 33 - Perfil dos usuários do turno da noite (das 18:00 até as 23:59 - A) e madrugada (das 00:00 até as 05:59 – B).



Fonte: do próprio autor.

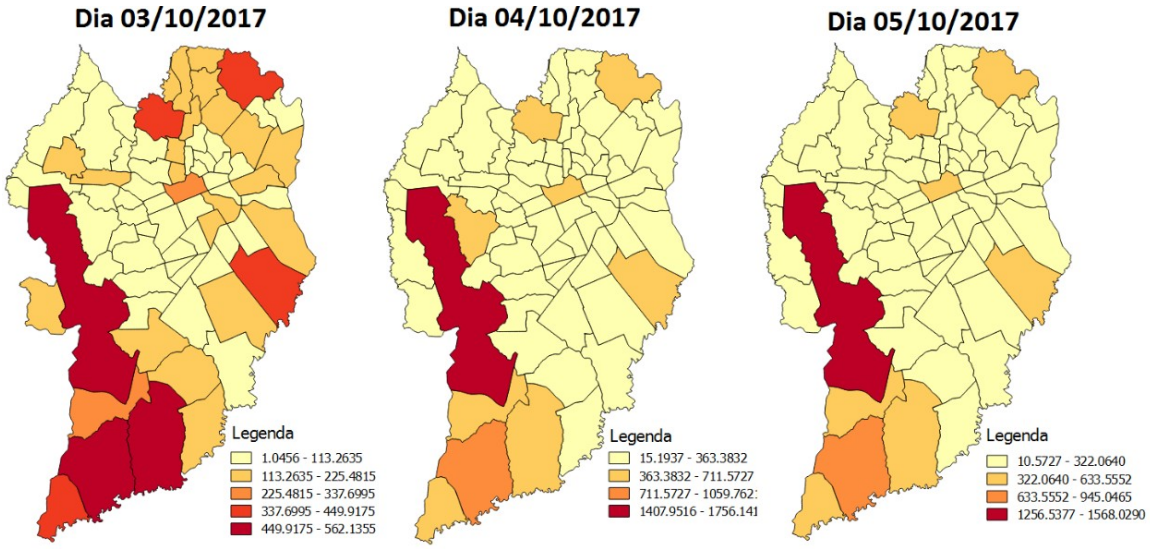
4.2 Embarques a nível de bairro

Em seguida, buscou-se verificar a quantidade de embarques diários sob o ponto de vista de bairro. Para tanto, buscou-se levar em consideração o Índice de Ocorrências relativo a densidade demográfica (*IOco*) calculada para cada bairro através da Equação 4.

$$IOco = \frac{quantidadeDeOcorrenciasNoBairro}{densidadeDemografica DoBairro} \quad (4)$$

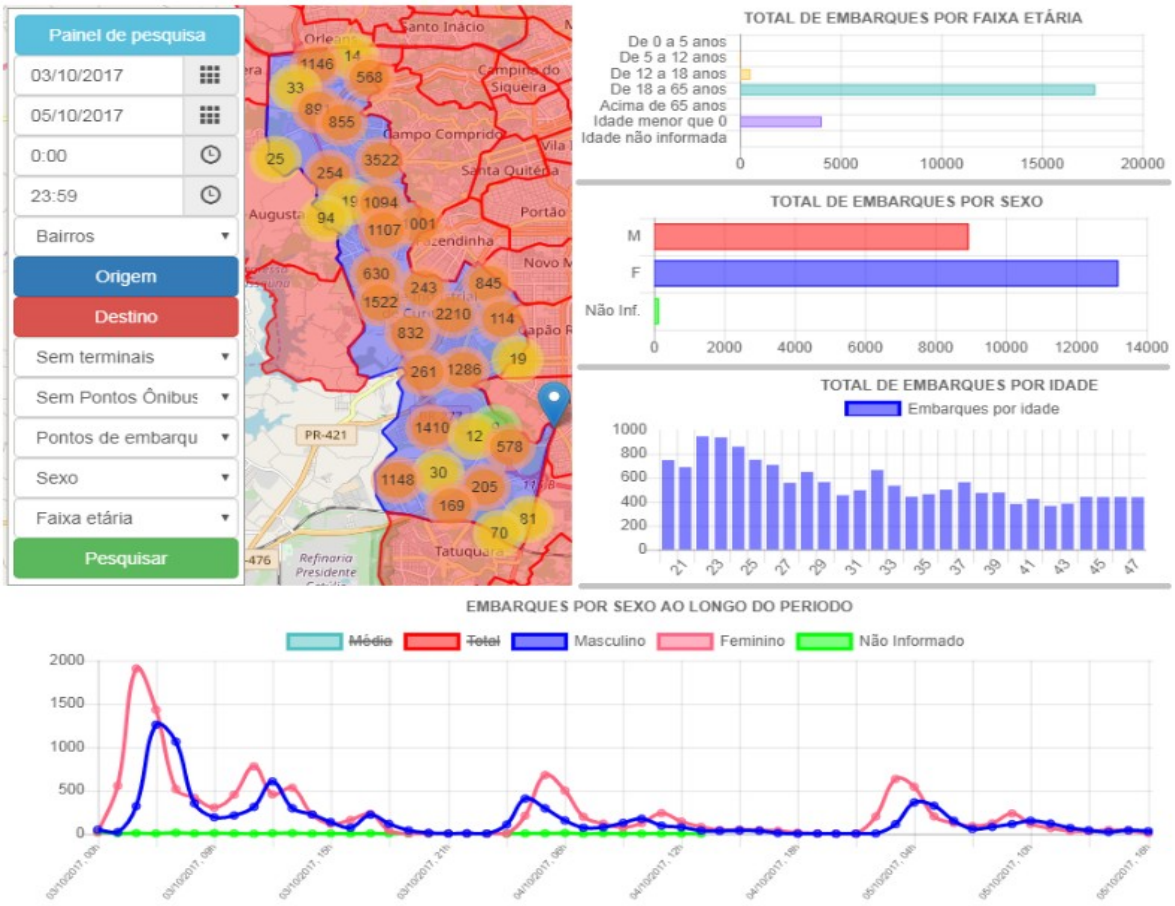
O resultado obtido pode ser visto na Figura 34: nota-se que o Cidade Industrial de Curitiba (CIC), bairro mais populoso daquela cidade (mais de 170 mil segundo senso do IBGE de 2010) e de maior área territorial da cidade (cerca de 44km^2) é também o de maior IOco (vide Equação 4), indicando que existe uma movimentação de pessoas consideravelmente maior que a densidade demográfica daquele bairro. Na sequência, recorreu-se ao nosso protótipo para investigar melhor tais embarques: para tanto, selecionou-se o CIC como bairro de origem e todos os demais bairros como destino. Os resultados obtidos são mostrados na Figura 35: verifica-se que a maior quantidade de embarques ocorre na região central daquele bairro. Com relação ao perfil dos usuários que de lá embarcam, constatou-se que são, na maioria, pessoas do sexo feminino com idade entre 18 e 65 anos, e cuja variação da demanda apresenta altos picos por volta das 06:00.

Figura 34 - Volume de embarques com base no *IOco*.



Fonte: do próprio autor.

Figura 35 - Perfil de embarques no CIC.

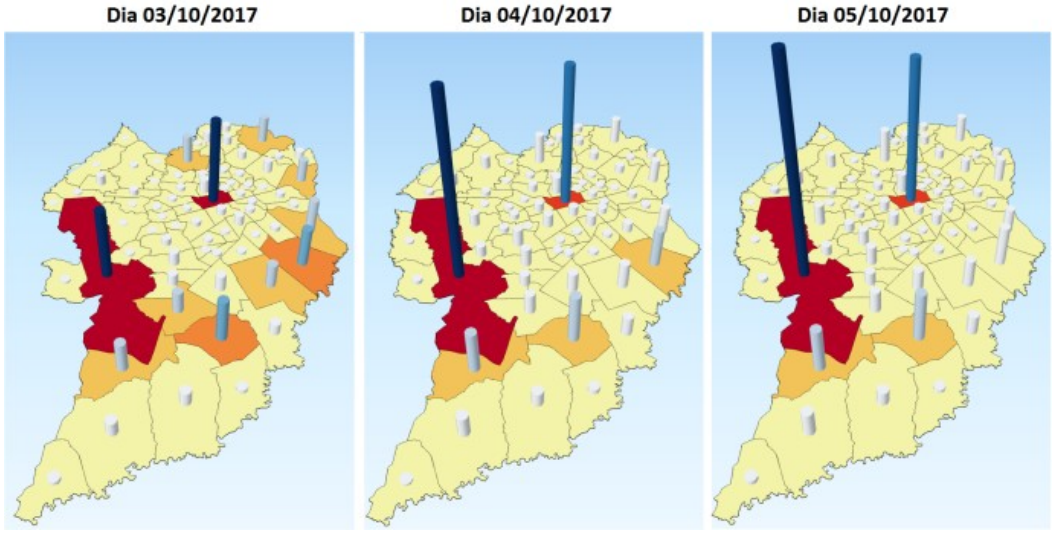


Fonte: do próprio autor.

4.3 Embarques a nível intrabairro

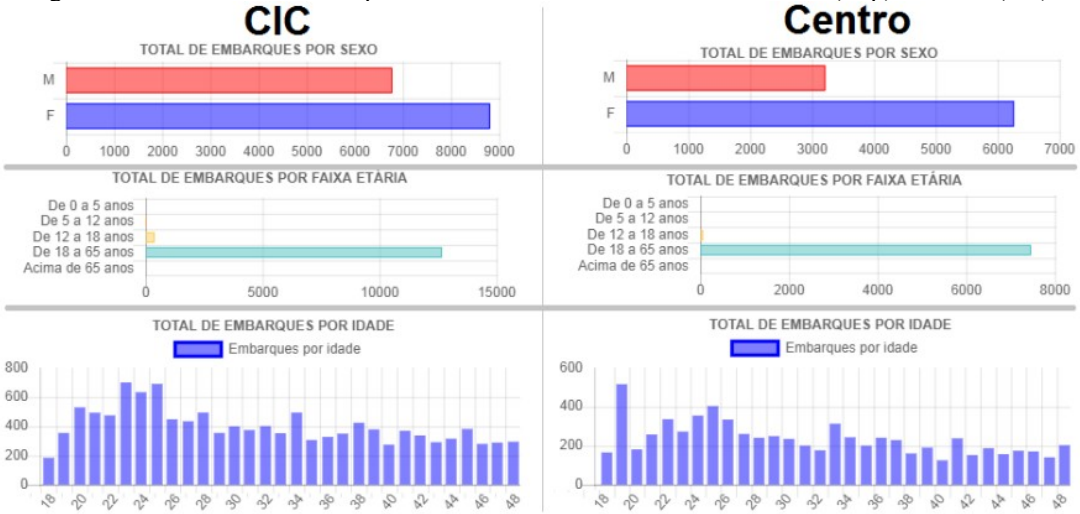
Também analisou-se o uso do STP de Curitiba do ponto de vista intrabairro, ou seja, cujos extremos (origem e destino) estivessem localizados dentro de um mesmo bairro. O resultado obtido é mostrado na Figura 36, onde é possível verificar que os bairros CIC e Centro são os bairros com a maior quantidade de ocorrências deste tipo. Nosso protótipo indicou que as pessoas que realizam tais deslocamentos são do sexo feminino e com idade entre 18 e 65 anos, conforme mostrada na Figura 37. No caso do bairro Centro, 2/3 dos usuários que realizam deslocamentos intrabairro são pessoas do sexo feminino.

Figura 36 - Deslocamentos intrabairro de Curitiba.



Fonte: do próprio autor.

Figura 37 - Perfil dos usuários que se deslocam no interior dos bairros CIC (esq.) e Centro (dir.).

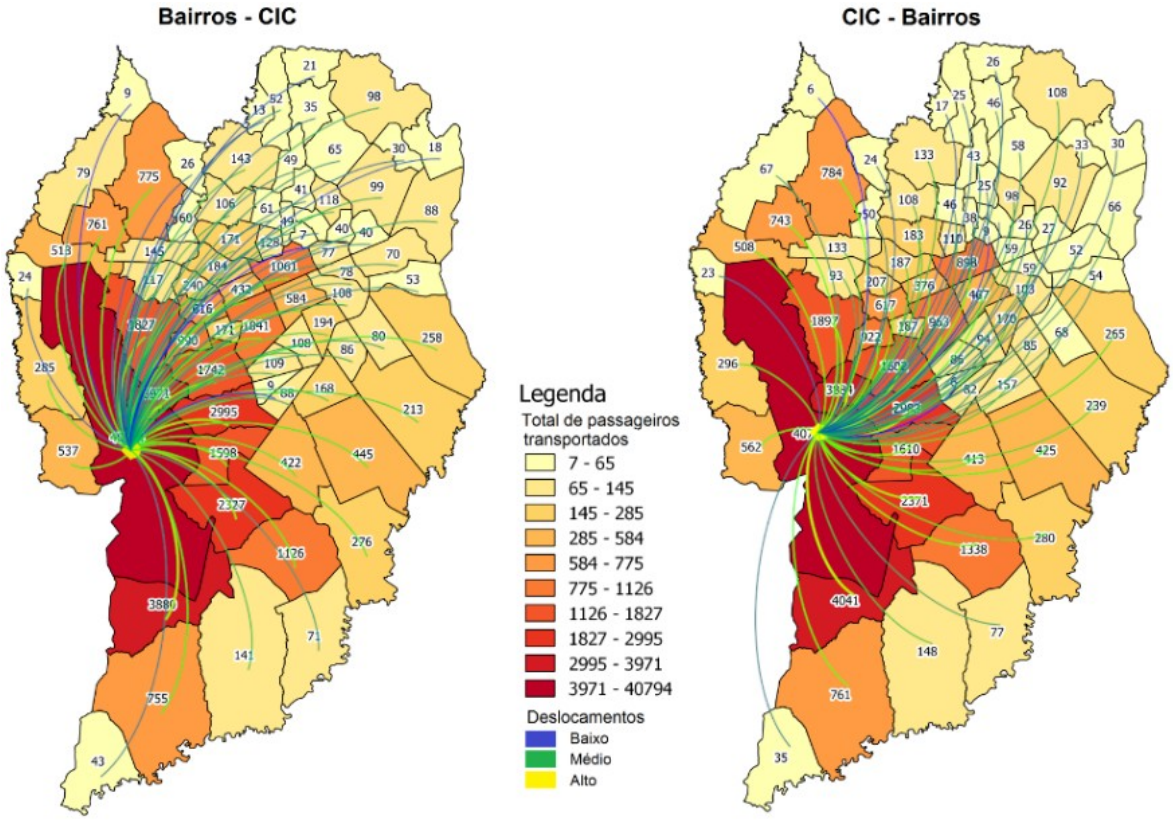


Fonte: do próprio autor.

4.4 Deslocamentos inter-bairros

Buscou-se também verificar a origem dos usuários que desembarcam no CIC e o destino dos que embarcam naquele bairro. A Figura 38 mostra o resultado obtido quando analisados 07 dias de dados (de 03/10 a 09/10/2017): quanto mais escuro o bairro, maior a quantidade de embarques/desembarques. Com relação as linhas, estas representam os deslocamentos dos usuários do STP. Assim, nota-se que os principais extremos de deslocamentos ocorridos ao longo de uma semana e que envolvem o CIC como um dos extremos são os bairros que o cercam. Uma explicação plausível seria que aqueles que trabalham e/ou estudam no CIC buscam morar naquele mesmo bairro ou em um próximo a ele e vice-versa.

Figura 38 - Deslocamentos com CIC como destino (esquerda) e como origem (direita).

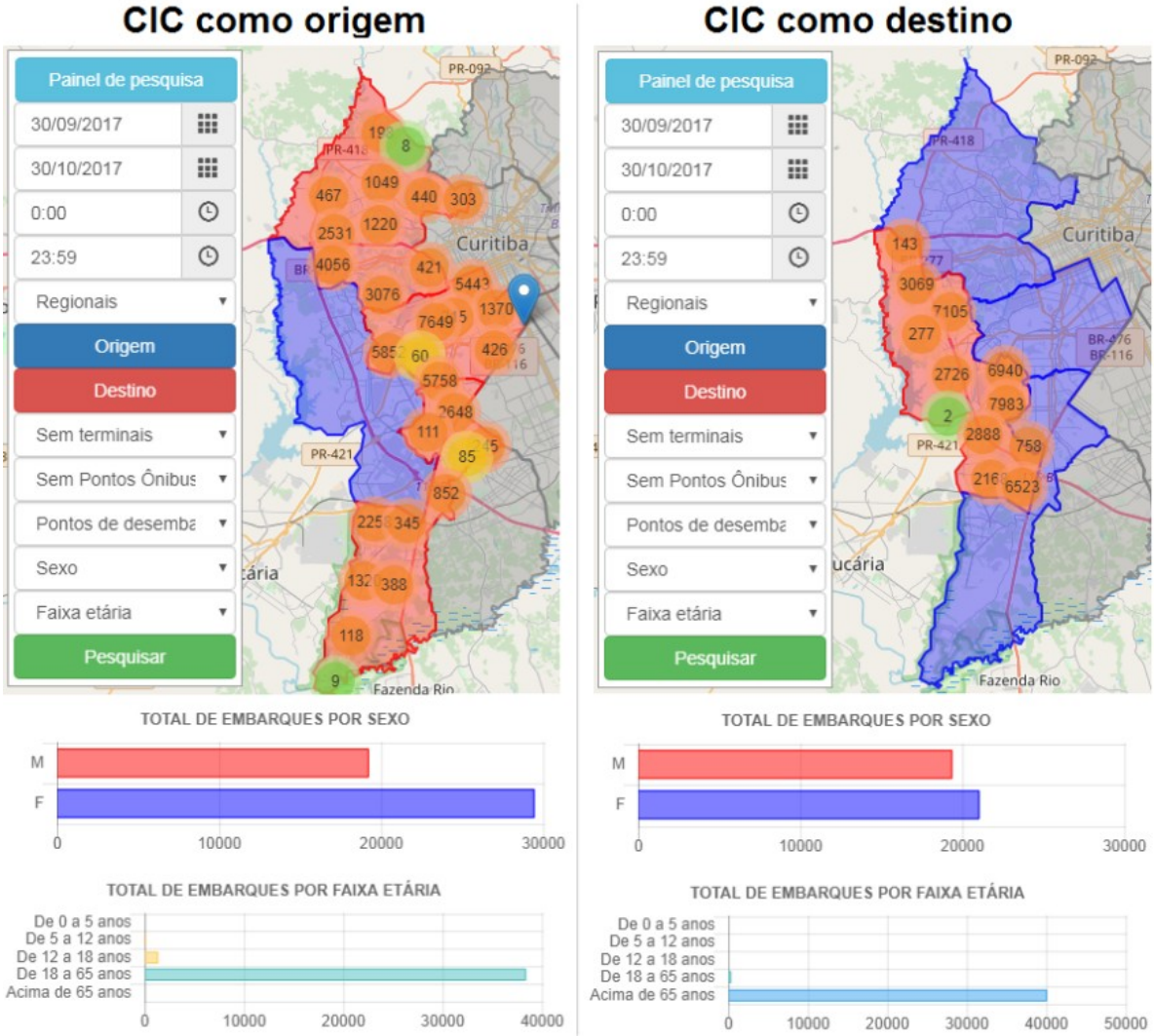


Fonte: do próprio autor.

Analisando-se os resultados mostrados na Figura 38, verificou-se que tais deslocamentos ocorrem entre os bairros pertencentes às regionais CIC, Santa Felicidade, Portão e Pinheirinho. Nesta direção, recorreu-se ao protótipo para obter maiores informações

acerca dos deslocamentos ocorridos entre tais regionais. Os resultados obtidos mostraram que a maioria dos que se deslocam da regional CIC são pessoas do sexo feminino com idade entre 18 e 65 anos com destinos às regionais Santa Felicidade ou Portão: vide Figura 39.

Figura 39 - Perfil dos deslocamentos entre as regionais CIC, Portão, Pinheirinho e Santa Felicidade.

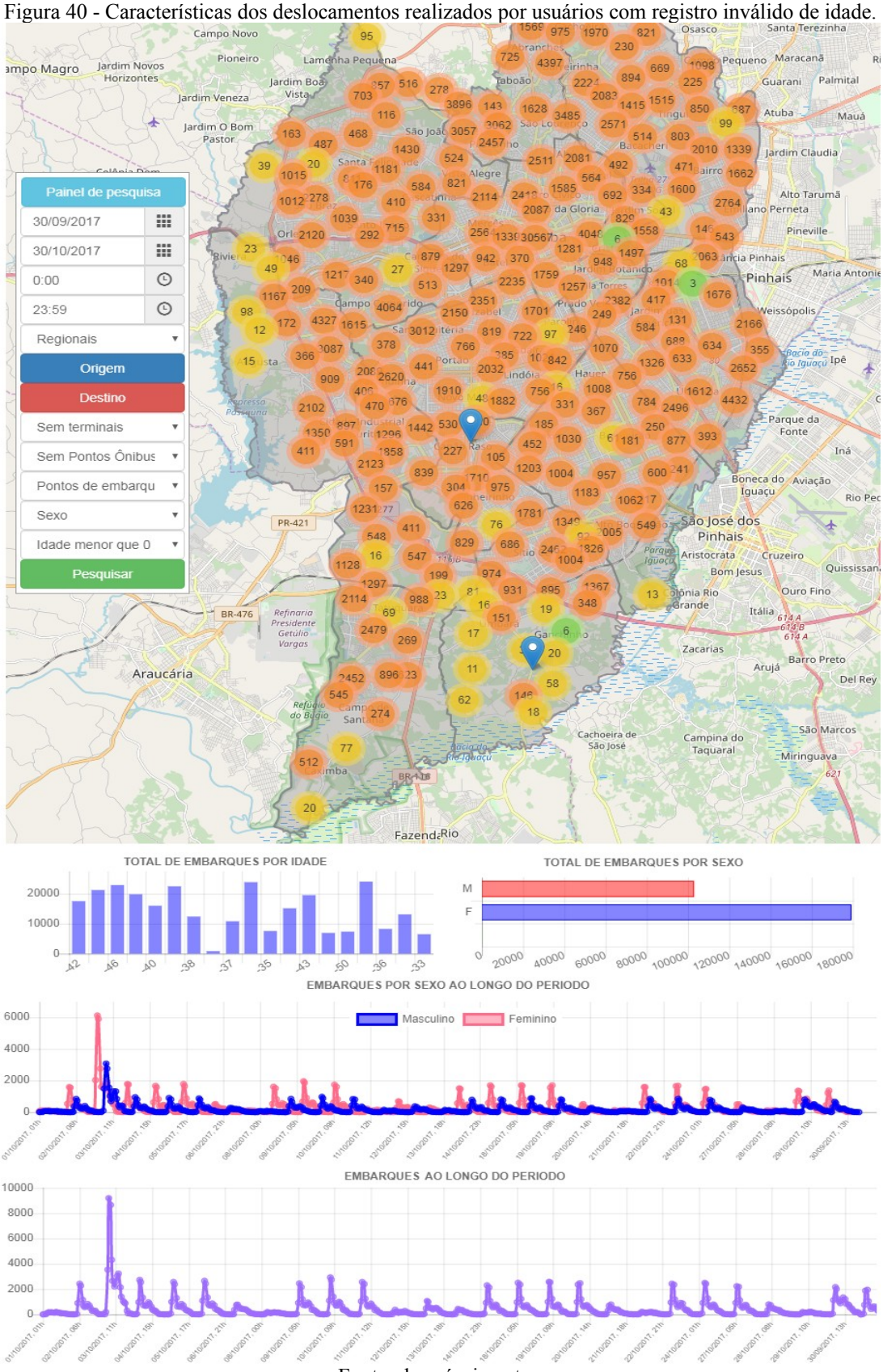


Fonte: do próprio autor.

4.5 Deslocamentos de usuários sem informação de idade

Durante o processo de desenvolvimento do protótipo, constatou-se que para muitos registros de usuários, o dado armazenado no campo idade era inválido. Nesta direção, nosso protótipo permitiu verificar que estes usuários utilizam com regularidade o STP de Curitiba, e costumam realizar embarques principalmente por volta das 06:00 dos dias úteis e são, na maioria, pessoas do sexo feminino (vide Figura 40).

Figura 40 - Características dos deslocamentos realizados por usuários com registro inválido de idade.



Fonte: do próprio autor.

5 CONCLUSÃO

Diversos são os desafios enfrentados principalmente pelos grandes centros urbanos relacionados à mobilidade urbana. A gestão da oferta e demanda do transporte público, o aumento da frota de veículos, os engarrafamentos cada vez mais frequentes, o aumento do tempo de viagens e a poluição ambiental são alguns dos problemas que impactam não só o funcionamento e desenvolvimento de uma cidade como também a qualidade de vida da população. Neste sentido, este trabalho apresentou uma nova ferramenta visual que suporta consultas espaço-temporais sobre dados de Origem-Destino de usuários do transporte público de Curitiba.

A primeira motivação surgiu durante a leitura de IPPUC (2017), o qual forneceu uma série de possibilidades de análises de dados de Origem-Destino: sob diferentes períodos, sob o ponto de vista de bairros e regionais de embarque e desembarque, e inclusive por sexo e faixa etária dos usuários. Outros trabalhos como os de Ferreira et al. (2011), Ferreira et al. (2013), Lu et al. (2015) e Zhang et al. (2015b) forneceram subsídios relativos a construção da interface de usuário, as possibilidades de filtro, as tecnologias utilizadas, a arquitetura, a forma de interação com o usuário e as diferentes possibilidades de aplicação de soluções deste tipo. Ainda, foram de grande valia as contribuições dadas por Vila (2016) no que tange as possibilidades de visualização e de clusterização de dados, e por Diniz Junior (2017) no que diz respeito aos métodos utilizados para dedução dos pontos de embarque e desembarque de usuários do STP de Curitiba. Um levantamento também foi realizado em meio a um grupo de pesquisadores da área de mobilidade urbana, o qual nos forneceu subsídios para decidirmos quais filtros oferecer no protótipo, como apresentar os dados, para quais fins a solução poderia ser utilizada, dentre outros.

O protótipo foi desenvolvido utilizando apenas tecnologias *open-source* e não demandou nenhum tipo de gasto referente a aquisição e/ou contratação de licença de *software*. A solução desenvolvida é inteiramente online e não exige a configuração ou instalação de nenhum tipo de *software* ou *plugin* para seu funcionamento. A ferramenta permite compreender aspectos relacionados ao perfil do usuário do transporte público, a variação da demanda do STP ao longo de um determinado período, os locais onde ocorrem a maior quantidade de embarques e desembarques, compreender a dinâmica de ocupação do espaço urbano, dentre outros. Resultados obtidos mediante realização de um teste de usabilidade

indicaram que a solução permite que usuários realizem de forma facilitada análises de OD sem precisar utilizar qualquer tipo de linguagem de programação e/ou de manipulação de dados.

A partir das análises realizadas, percebeu-se que a demanda geral do STP de Curitiba se mantém relativamente constante durante os dias úteis, sofrendo uma diminuição nos finais de semana. O pico de utilização do STP daquela cidade é o mesmo para dias úteis e não-úteis, concentrando-se no turno da manhã e da tarde. Do ponto de vista de deslocamentos, o CIC é o bairro onde ocorre a maior quantidade de embarques e desembarques. É também no CIC onde ocorre a maior quantidade de deslocamentos a nível intrabairro, ou seja, com extremos (origem e destino) dentro do próprio bairro. Tal fato ocorre inclusive quando analisado tais dados levando-se em consideração a densidade demográfica dos bairros, o que evidencia a importância do bairro Cidade Industrial de Curitiba em termos de mobilidade urbana para aquela cidade.

A realização deste trabalho evidenciou uma série de desafios relacionados ao desenvolvimento de uma solução de visualização de dados de Origem-Destino, tais como buscar definir o quanto antes o público-alvo da solução, seu nível de entendimento no assunto em questão, sua experiência no uso das tecnologias envolvidas, e quais funcionalidades deseja encontrar na ferramenta a ser desenvolvida. A partir daí, o desafio passa a ser encontrar tecnologias e metodologias que permitam desenvolver uma solução que atenda aos anseios do público-alvo e que também permita entregar resultados não só relevantes, mas também na forma e densidade compatíveis com a capacidade de compreensão do usuário. Neste cenário, a inevitável necessidade de integração de diferentes tecnologias pode conduzir a impasses decorrentes de incompatibilidade de recursos. Também é importante verificar o tempo de resposta da aplicação face ao volume de dados envolvido.

Dentre os possíveis trabalhos futuros identificados, sugere-se 1) aperfeiçoar a estimativa de destino de deslocamentos, 2) aperfeiçoar a interface da ferramenta, 3) ampliar a oferta de filtros de pesquisa (taxa de ocupação dos veículos, velocidade média dos deslocamentos, etc), 4) permitir que o usuário possa desenhar no mapa as regiões a partir das quais deseja efetuar análises de OD, 5) permitir a realização de consultas baseando-se não apenas em regiões de interesse, mas também em pontos de interesse, 6) otimizar o tempo das consultas em geral, 7) automatizar a carga de dados na aplicação, 8) permitir a adição de novas camadas ao mapa (linhas de fluxo de deslocamentos, alertas de velocidade, de desvio de rota e de paralisação de operação, dados relativos a fatores climáticos, etc).

REFERÊNCIAS BIBLIOGRÁFICAS

ALDAMA-NALDA, A.; CHOURABI, H.; PARDO, T. A.; GIL-GARCIA, J. R.; MELLOULI, S.; SCHOLL, H. J.; ALAWADHI, S.; NAM, T.; e WALKER, S. Smart cities and service integration initiatives in north american cities: a status report. *In:* INTERNATIONAL CONFERENCE ON DIGITAL GOVERNMENT RESEARCH, 13., 2012, USA. **Anais** […]. New York: ACM. 2012. p. 289–290.

ANDRADE, T. C.; PEREIRA, M. A.; WANNER, E. F. Development of an application using a clustering algorithm for definition of collective transportation routes and times. *In:* BRAZILIAN SIMPOSIUM ON GEOINFORMATICS, 19., 2014, São Paulo. **Anais** […], São Paulo: SBC. 2012. p. 13–24.

AÑEZ, J.; BARRA, T. D. L.; PÉREZ, B. Dual graph representation of transport networks. **Transportation Research Part B: Methodological**, v. 3, n. 3, p. 209-216, Elsevier, 1996.

ASSOCIAÇÃO NACIONAL DE TRANSPORTES PÚBLICOS. **Transporte humano**: cidades com qualidade de vida. 1. ed. ANTP. 1997. 312 p.

BARCZYSZYN, G. L. **Integração de dados geográficos para planejamento urbano da cidade de Curitiba**. 2015. 88 f. Trabalho de Conclusão de Curso (Graduação), Universidade Tecnológica Federal do Paraná, Paraná, 2015.

BELUZO, J. R. **Ideo: Integrador de dados da execução orçamentária brasileira**. 2015. 126 f. Dissertação (Mestrado em Ciências) - Universidade de São Paulo, São Paulo, 2015.

BORBA, E. M. **Medidas de centralidade em grafos e aplicações em redes de dados**. 2013. 61 f. Dissertação (Mestrado em Matemática) - Universidade Federal do Rio Grande do Sul, Rio Grande do Sul, 2013.

CARD, M. **Readings in information visualization**: using vision to think. Morgan Kaufmann, 1999. 712 p.

CASSIANO, K. M. **Análise de séries temporais usando análise espectral singular (ssa) e clusterização de suas componentes baseada em densidade**. 2014. 172 f. Tese (Doutorado em Engenharia Elétrica) - Pontifícia Universidade Católica, Rio de Janeiro. 2014.

CHAPLEAU, R.; MORENCY, C. Dynamic spatial analysis of urban travel survey data using GIS. *In:* ANUAL ESRI INTERNATIONAL USER CONFERENCE, 25., 2005, California. **Anais** […]. California: ESRI, 2005, p. 1–14.

COLE, R. M. **Clustering with genetic algorithms**. 1998. 110 f. Dissertação (Mestrado em Ciências) - University of Western Australia, Australia. 1998.

COSTA, G.; KOZIEVITCH, N. P.; FONSECA, K.; GADDA, T.; BERARDI, R. Integração de dados de redutores de velocidade no transporte público de Curitiba. *In:* ESCOLA

REGIONAL DE BANCO DE DADOS, 9., 2017, Rio Grande do Sul. **Anais** […]. Rio Grande do Sul: SBC. 2017. p. 123–126.

COUTO, E. A. **Aplicação dos indicadores de desenvolvimento sustentável da norma ABNT NBR ISO 37120:2017 para a cidade do Rio de Janeiro e análise comparativa com cidades da América Latina**. 2018. 163 p. Trabalho de Conclusão de Curso (Graduação) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2018.

CRUZ, A.; FERREIRA, J.; CARVALHO, D.; MENDES, E.; PACITTI, E.; COUTINHO, R.; Porto, F.; OGASAWARA, E. Detecção de anomalias frequentes no transporte rodoviário urbano. *In:* BRAZILIAN SYMPOSIUM ON DATABASES, 22., 2018, Rio de Janeiro. **Anais** […]. Rio de Janeiro: SBC. 2018. p. 271–276.

MATTOS, V. G.; VASCONCELOS, P. H.; PARCIANELLO, Y.; KOZIEVITCH, N. P.; BERARDI, R. Visualização dos dados abertos da polícia rodoviária federal sobre acidentes nas rodovias brasileiras. *In:* SIMPÓSIO BRASILEIRO DE BANCO DE DADOS, 34., 2019, Ceará. **Anais** […]. Ceará: SBC, 2019, p. 193–198.

DIAS, J.; da SILVA, T.; MERGULHÃO, R.; VIEIRA, J. Proposta de agrupamento das cidades médias brasileiras para elaboração do plano de mobilidade urbana. II Workshop de Computação Social. 2016.

DINIZ JUNIOR, P. C. **Serviços telemáticos em uma rede de transporte público baseados em veículos conectados e dados abertos**. 2017. 114 f. Dissertação (Mestrado em Computação Aplicada) - Universidade Tecnológica Federal do Paraná, Paraná, 2017.

FERREIRA, N.; LINS, L.; FINK, D.; KELLING, S.; WOOD, C.; FREIRE, J.; SILVA, C. Birdvis: Visualizing and understanding bird populations. **IEEE Transactions on Visualization and Computer Graphics**, v. 17, n. 12, p. 2374-2383, Elsevier, 2011.

FERREIRA, N.; POCO, J.; VO, H. T.; FREIRE, J.; SILVA, C. T. Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips. **IEEE Transactions on Visualization and Computer Graphics**, v. 19, n. 12, p. 2149-2158, IEEE, 2013.

FREITAS, C. M. D. S.; CHUBACHI, O. M.; LUZZARDI, P. R. G.; CAVA, R. A. Introdução à visualização de informações. **Revista de informática teórica e aplicada**. Porto Alegre. v. 8, n. 2, p. 143-158, 2001.

GAY, J. S.; GIANNOTTI, M. A.; TOMASIELLO, D. B. Accessibility and flood risk spatial indicators as measures of vulnerability. **Revista Brasileira de Cartografia**, v. 69, n. 5, 2018.

GUERRA, A. L. **Determinação de matriz origem destino utilizando dados do sistema de bilhetagem eletrônica**. 2011. 116 f. Dissertação (Mestrado em Geotecnia e Transportes) - Universidade Federal de Minas Gerais, Minas Gerais, 2011.

GUERRA, A. L.; BARBOSA H. M.; DE OLIVEIRA, L. K. Estimativa de matriz origem-destino utilizando dados do sistema de bilhetagem eletrônica: proposta metodológica. **Transportes**, v. 22, n. 3, p. 26–38. 2014.

HARRIS, M. **How a lone hacker shredded the myth of crowdsourcing.** Disponível em: https://www.wired.com/2015/02/how-a-lone-hacker-shredded-the-myth-of-crowdsourcing/. Acesso em: Mai. 30, 2018.

HARTIGAN, J. A. **Clustering Algorithms**. 1. ed. Wiley, 1975. 351 p.

IPPUC. **Consolidação de dados de oferta, demanda, sistema viário e zoneamento. relatório 5**: Pesquisa de origem-destino domiciliar. Disponível em: https://ippuc.org.br/visualizar.php?doc=http://admsite2013.ippuc.org.br/arquivos/documentos/D536/D536_002_BR.PDF. Acesso em: Dez. 10, 2017.

JERÔNIMO, C. L. M.; CAMPELO, C. E. C.; BAPTISTA, C. S. Analyzing mobility patterns from social networks and social, economic and demographic Open Data. *In:* Proceedings of BRAZILIAN SYMPOSIUM ON GEOINFORMATICS, 17., 2016, São Paulo. **Anais** […]. São Paulo: SBC, 2016. p. 32–43.

KONO, F. A. M. **Um modelo de representação computacional baseado em conceitos de crescimento urbano associados a alvarás e primitivas em banco de dados espacial**. 2016. 143 f. Dissertação (Mestrado em Computação Aplicada) - Universidade Tecnológica Federal do Paraná, Paraná, 2016.

KOZIEVITCH, N. P.; ALMEIDA, L. D. A.; SILVA, R. D.; MINETTO, R. An alternative and smarter route planner for wheelchair users: exploring open data. *In:* INTERNATIONAL CONFERENCE ON SMART CITIES AND GREEN ICT SYSTEMS, 5., 2016, Itália. **Anais** […]. Itália: IEEE, 2016, p. 1–6.

KOZIEVITCH, N. P.; PARCIANELLO, Y.; FONSECA, K. V. O.; ROSA, M. O.; GADDA, T. M. C.; MALUCELLI, F. C. Transportation: An overview from Open Data approach. *In:* INTERNATIONAL SMART CITIES CONFERENCE, 4., 2018, USA. **Anais** […]. USA: IEEE, 2018, p. 1–8.

KOZIEVITCH, N. P.; SILVA, T. H.; ZIVIANI, A.; COSTA, G.; e LUGO, G. Three decades of business activity evolution in Curitiba: a case of study. **Annals of Data Science**, v. 4, n. 3, p. 307–327, Springer, 2017.

LI WEIGANG, K., W.; YAMASHITA, Y.; MACIVER, A. Algorithms for estimating bus arrival times using gps data. *In:* INTERNATIONAL CONFERENCE ON INTELLIGENT TRANSPORTATION SYSTEMS, 5., 2002, Singapore. **Anais** […]. Singapore: IEEE, 2002, p. 868–873.

LU, M.; WANG, Z.; LIANG, J.; YUAN, X. Od-wheel: Visual design to explore od patterns of a central region. *In:* VISUALIZATION SYMPOSIUM, 1., 2015, China. **Anais** […]. China: IEEE, 2015, p. 87–91.

MINETTO, R.; KOZIEVITCH, N. P.; SILVA, R. D.; ALMEIDA, L. D. A.; , SANTI, J. Shortcut suggestion based on collaborative user feedback for suitable wheelchair route planning. *In:* INTERNATIONAL CONFERENCE ON INTELLIGENT TRANSPORTATION SYSTEMS, 19., 2016, Rio de Janeiro. **Anais** […]. Rio de Janeiro: IEEE, 2016, p. 2372-2377.

NATIONS, U. **Population facts: the speed of urbanization around the world.** Disponível em: https://www.un.org/en/development/desa/population/publications/PDF/popfacts/PopFacts_2018-1.PDF. Acesso em: Jun. 07, 2018.

OSAMA, D.; GHONEIM, A.; MANJAUNATH, B. Air pollution clustering using k means algorithm in smart city. **International Journal of Innovative Research in Computer and Communication Engineering.** v. 3, n. 7, p. 51–57, IJIRCCE, 2015.

PARCIANELLO, Y; KOZIEVITCH, N. P. **Exploratory analysis of public daily expenses from the government of Rio Grande do Sul**. II Workshop de Computação Social. 2017.

RAMOS, C. S. **Visualização cartográfica e cartografia multimídia.** 1. ed. UNESP, 2005. 184 p.

RODRIGUE, J. P. **The Geography of Transport Systems**. 4. ed. Routledge, 2017. 454 p.

RODRIGUES, M. Introdução ao geoprocessamento. *In:* Simpósio Brasileiro de Geoprocessamento, volume 1, p. 1–26. São Paulo. 1990.

SABERI, M.; MAHMASSANI, H. S.; BROCKMANN, D.; HOSSEINI, A. (2017). A complex network perspective for characterizing urban travel demand patterns: graph theoretical analysis of large-scale origin–destination demand networks. **Transportation**, v. 44, n. 6, p. 1383-1402, Springer, 2017.

SABIDUSSI, G. The centrality index of a graph. **Psychometrika**, v. 31, n. 4, p. 581–603, Springer, 1996.

SILVA, E.; ROSA, M.; FONSECA, K.; LUDERS, R.; KOZIEVITCH, N. Combining k-means Method and complex network analysis to evaluate city mobility. *In:* INTELLIGENT TRANSPORTATION SYSTEMS, 19., 2016, Rio de Janeiro. **Anais** […]. Rio de Janeiro: IEEE, 2016, p. 1666–1671

SIMETTE, G.; PARCIANELLO, Y.; KOZIEVITCH, N. P.; FONSECA, K. V. O. Análise da situação dos redutores de velocidade de Curitiba. *In:* ESCOLA REGIONAL DE BANCO DE DADOS, 14., Rio Grande do Sul. **Anais** […]. Rio Grande do Sul: SBC, 2018, p. 123–126.

SOARES, A. C. **Diagnóstico e modelagem da rede de distribuição de derivados de petróleo no Brasil**. 2003. 156 f. Dissertação (Mestrado em Geotecnia e Transportes) - Pontifícia Universidade Católica, Rio de Janeiro, 2003.

SOUZA, R.; OLIVEIRA, I. P.; SALES, L.; FERRAZ, F. Beyond efficiency: how to use geolocation applications to improve citizens well-being. *In:* INTERNATIONAL CONFERENCE ON SMART SYSTEMS, DEVICES AND TECHNOLOGIES, 4., 2017, Belgium. **Anais** […]. Belgium: IARA, 2017, p. 37–40.

SPADON, G.; SCABORA, L. C.; NESSO, Jr.; M. R.; TRAINA, Jr.; C.; RODRIGUES, Jr.; J. F. Caracterização topológica de redes viárias por meio da análise de vetores de características e técnicas de agrupamento. *In:* BRAZILIAN SYMPOSIUM ON DATABASES, 33., 2018. **Anais** […]. Rio de Janeiro: SBC. 2018, p. 157–168.

STOLFI, D. H.; ALBA, E.; YAO, X. Predicting car park occupancy rates in smart cities. *In:* INTERNATIONAL CONFERENCE ON SMART CITIES, 1., 2017, Algeria. **Anais** […]. Algeria: Springer. 2017, p. 107–117.

TIBSHIRANI, R.; WALTHER, G.; HASTIE, T. Estimating the number of clusters in a data set via the gap statistic. **Journal of the Royal Statistical Society**, v. 63, n. 2, p. 411-423, Royal Statistical Society, 2001.

UNICEF. **Uprooted: the growing crisis for refugee and migrant children**. Disponível em: https://www.unicef.org/publications/index_92710.html. Acesso em: Mar. 15, 2017.

VILA, J. J. F. R. **Clusterização e visualização espaço-temporal de dados georreferenciados adaptando o algoritmo marker clusterer: um caso de uso em Curitiba**. 2016. 87 f. Dissertação (Mestrado em Computação Aplicada) - Universidade Tecnológica Federal do Paraná, Paraná, 2016.

VILA, J. J. R.; KOZIEVITCH, N. P.; GADDA, T. M. C.; FONSECA, K. V. O.; ROSA, M. O.; GOMES JUNIOR, L. C.; AKBAR, M. Urban mobility challenges: an exploratory analysis of public transportation data in Curitiba. **Revista de Informática Aplicada**, São Paulo, v. 12, p. 1-14, 2016.

VON FERBER, C.; HOLOVATCH, T.; HOLOVATCH, Y.; PALCHYKOV, V. Public transport networks: empirical analysis and modeling. **The European Physical Journal B**, v. 68, n. 2, p. 261-275, Springer, 2009.

WASSERMAN, S.; FAUST, K. **Social network analysis**: Methods and applications. 8. ed., Cambridge University Press, 1994. 868 p.

WORBOYS, M. F.; DUCKHAM, M. **GIS: a computing perspective**. 2. ed., CRC Press, 2004. 448 p.

YU, C. H. **Exploratory data analysis**. 1. ed. Pearson**,** 1977. 688 p.

ZAIANE, O. R.; FOSS, A.; LEE, C. H.; WANG, W. On data clustering analysis: Scalability, constraints, and validation. *In:* PACIFIC-ASIA CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 1., 2002, Berlin. **Anais** […]. Berlin: Springer, 2002, p. 28–39.

ZHANG, H.; ZHAO, P.; WANG, Y.; YAO, X.; ZHUGE, C. Evaluation of bus networks in china: from topology and transfer perspectives. **Discrete Dynamics in Nature and Society**, v. 2015, p. 1–8, 2015a.

ZHANG, J.; YOU, S.; XIA, Y. Prototyping a web-based high performance visual analytics platform for origin-destination data: A case study of nyc taxi trip records. *In:* INTERNATIONAL ACM SIGSPATIAL WORKSHOP ON SMART CITIES AND URBAN ANALYTICS, 1., 2015, USA. **Anais** […]. USA: ACM. 2015b, p. 16–23.

APÊNDICE A - FORMULÁRIO DE PESQUISA PRÉ-PROTÓTIPO

Quiz

The responses of this quiz are anonymous and will be used to help developing of a prototype related to a master's degree project from UTFPR PPGCA. Thank you for your cooperation.

1. **Do you use public transportation?**
   Marcar apenas uma oval.
   - ◯ No
   - ◯ Daily
   - ◯ Only monday to friday
   - ◯ Only on weekends
   - ◯ Eventually

2. **Have you ever had contact with any type of work involving the analysis of the use of public transportation in Curitiba?**
   Marcar apenas uma oval.
   - ◯ Yes ◯ No

3. **In your opinion, would it be interesting to have a solution that allows you to visualize, quantify and exploit data related to use of public transportation in Curitiba?**
   Marcar apenas uma oval.
   - ◯ Yes ◯ No

4. **In this solution, what types of search filters would you like to find?**
   Marque todas que se aplicam.
   - ☐ Filter trips by boarding region
   - ☐ Filter trips by landing region
   - ☐ Filter trips by time interval
   - ☐ Filter trips by users' age range
   - ☐ Filter trips by users' gender
   - ☐ Outro: _____

5. **For you, what would be the possibilities of using this solution?**
   Marque todas que se aplicam.
   - ☐ Study of the demand of public transportation
   - ☐ Analysis of the profile of public transportation users
   - ☐ Study of the dynamics of urban space occupation
   - ☐ Outro: _____

6. **What other kind of suggestion would you give for this app which would serve to visualizes data from buses?**

   _____

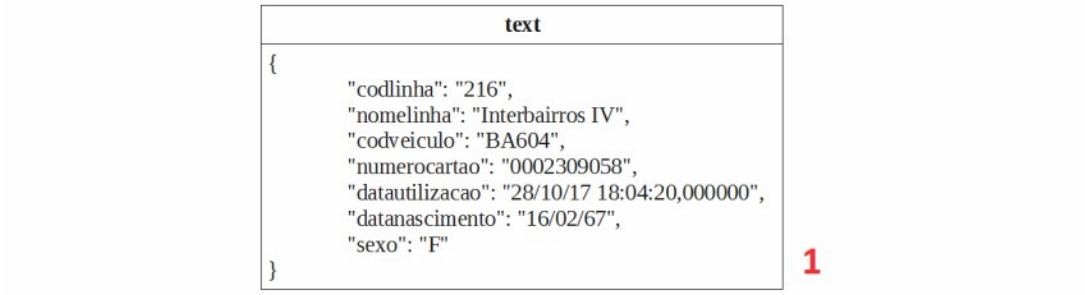
   _____

   _____

APÊNDICE B - INFORMAÇÕES SOBRE A CARGA DOS DADOS

A Figura 41 permite visualizar a diferença da estrutura de um registro de uso de cartão (conteúdo em formato *JSON* disponibilizado em arquivo com extensão *TXT*) e após a importação.

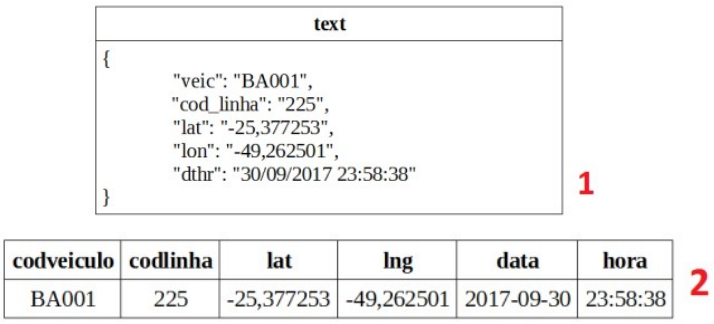
Figura 41 - Dados de cartões antes (1) e após a importação (2).



Fonte: do próprio autor.

Procedeu-se de forma análoga para realizar a importação dos dados referentes a posição dos veículos para dentro da tabela "onibus". A Figura 42 permite visualizar a diferença da estrutura de um registro de posição de ônibus (conteúdo em formato *JSON* disponibilizado em arquivo com extensão *TXT*) e após a importação.

Figura 42 - Dados de posições dos veículos antes(1) e após a importação (2).



Fonte: do próprio autor.

A Tabela 4 traz algumas informações relacionadas ao tempo gasto na operação de importação dos dados para dentro da base no Postgre*SQL*.

Tabela 4 - Tempo médio de importação dos arquivos para a base de dados.

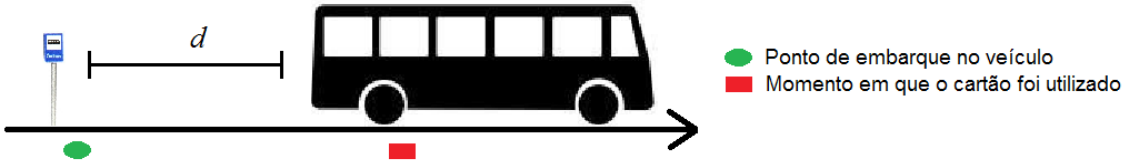
Termo contabilizado	Dados cartões	Posição dos veículos
Quantidade total de registros	8.710.082	114.602.481
Tempo médio de importação por arquivo *TXT*	±1 min, cada	±10 min, cada
Tempo total de importação (31 arquivos *TXT*)	±30 min	±5 hr

APÊNDICE C - DEDUÇÃO DOS PONTOS DE EMBARQUE

Optou-se por adotar a abordagem utilizada por Diniz Junior (2017) para deduzir o ponto aproximado de embarque dos usuários do STP: considerar como tal a posição do veículo no qual este usuário embarcou imediatamente após o instante do uso do seu respectivo cartão. Para tanto, foi elaborada uma sentença *SQL* para realizar uma operação de *join* das tabelas de cartões e de posição de veículos. Índices foram providenciados para agilizar o tempo de resposta deste procedimento.

Cabe ressaltar que a posição de embarque do usuário e a posição do veículo no momento em que o passageiro utiliza o cartão tendem a não ser as mesmas, conforme ilustra a Figura 43. Apesar de a abordagem aqui adotada não permitir encontrar o ponto exato de embarque dos passageiros, ainda assim é uma aproximação válida para os fins pretendidos por esta pesquisa.

Figura 43 - Ponto de embarque *vs* ponto de uso do cartão.



Fonte: do próprio autor.

Uma série de inconsistências foram identificadas durante o processo de definição de pontos de embarque dos passageiros. Percebeu-se que para muitos registros de cartões de usuários, não foram encontradas correspondências no *dataset* das posições de veículos. Em relação aos veículos, muitos dos registros possuíam os campos latitude/longitude vazios ou fora dos limites de Curitiba. Tais casos foram removidos do *dataset*. Os números referentes aos dados resultantes desta operação são mostrados na Tabela 5.

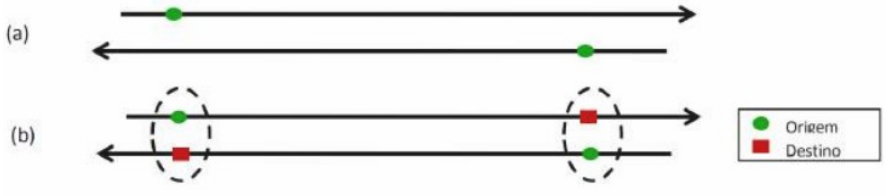
Tabela 5 - Números do processo de definição de pontos de embarque.

Questão norteadora	Número de registros
Quantos registros de uso de cartão dispúnhamos inicialmente?	8.710.082
Para quantos destes foi possível deduzir o ponto de embarque?	5.743.791

APÊNDICE D - DEDUÇÃO DOS PONTOS DE DESEMBARQUE

Como os *dataset*s analisados não continham o local de desembarque dos passageiros, valeu-se mais uma vez da abordagem utilizada por Diniz Junior (2017), convencionando-se como ponto de desembarque de uma viagem o ponto de embarque da viagem imediatamente posterior. A Figura 44 ilustra o processo adotado nesta pesquisa. A Tabela 6 mostra os números referentes a este processo.

Figura 44 - Processo de definição dos pontos de desembarque.



Fonte: Guerra et al. (2014).

Tabela 6 - Números do processo de definição de pontos de desembarques.

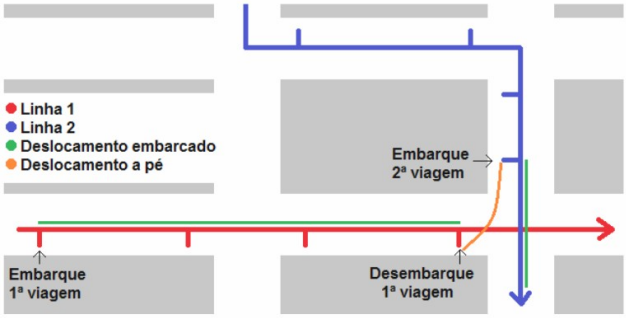
Situação	Qtd
Quantos registros dispúnhamos sem ponto de desembarque?	5.743.791
Para quantos destes foi possível deduzir o ponto de desembarque?	1.378.733

Vale destacar que na metodologia utilizada para definir os pontos de desembarque dos usuários do STP de Curitiba, apresenta limitações:

• Não foi possível estimar os destinos de usuários que haviam utilizado seus cartões apenas 01 vez por dia durante o período analisado. As tuplas referentes a tais situações foram excluídas do *dataset*.

• Desconsiderou-se a possibilidade de um usuário ter se deslocado (a pé, de táxi, bicicleta...) desde o ponto onde desembarcou do STP até o ponto onde tornou a embarcar. A Figura 45 ilustra esta situação.

Figura 45 - Ponto de desembarque estimado *vs* real.



Fonte: do próprio autor.

APÊNDICE E - DICIONÁRIO DE DADOS

Tabela 7 - Dicionário de dados da base do protótipo.

Tabela	Campo	Descrição
ROIs	id	Identificador da tupla
	lat	Coordenada de latitude (SRID 4326)
	lng	Coordenada de longitude (SRID 4326)
	geom	Geometria (SRID 29192)
pontos_onibus	id	Identificador da tupla
	lat	Coordenada de latitude (SRID 4326)
	lng	Coordenada de longitude (SRID 4326)
	geom	Geometria (SRID 29192)
terminais	id	Identificador da tupla
	nome	Descrição do terminal
	lat	Coordenada de latitude (SRID 4326)
	lng	Coordenada de longitude (SRID 4326)
	geom	Geometria (SRID 29192)
movimentacao	codlinha	Código da linha operada pelo ônibus
	nomelinha	Nome da linha operada pelo ônibus
	codveiculo	Código do veículo em operação
	cartao_numero	Número do cartão do passageiro
	cartao_datahora	Data-hora do uso do cartão do passageiro
	cartao_data	Data do uso do cartão do passageiro
	cartao_hora	Hora do uso do cartão do passageiro
	cartao_data_nascimento	Data de nascimento do passageiro
	cartao_sexo	Sexo do passageiro
	origem_lat	Coordenada de latitude da origem (SRID 4326)
	origem_lng	Coordenada de longitude da origem (SRID 4326)
	origem_datahora	Data-hora do embarque
	destino_lat	Coordenada de latitude do destino (SRID 4326)
	destino_lng	Coordenada de longitude do destino (SRID 4326)
	destino_datahora	Data-hora do desembarque
	origem	Geometria do ponto de origem (SRID 29192)
	destino	Geometria do ponto de destino (SRID 29192)
	idade	Idade do passageiro

APÊNDICE F - SENTENÇAS *SQL* UTILIZADAS PARA OTIMIZAÇÃO

Consulta A

```
select movimentacoes.cartao_sexo, movimentacoes.idade, movimentacoes.cartao_data ,
to_char (movimentacoes.cartao_datahora , 'DD/MM/YYYY, HH24h') as datahora_formatada,
movimentacoes.origem_lat, movimentacoes.origem_lng, movimentacoes. destino_lat , movimentacoes.destino_lng
from
( select id , ROI as contornoPg from transporte_dinamico . yp_ROIs where id in (2029)) origens ,
( select id , ROI as contornoPg from transporte_dinamico . yp_ROIs where id in (2013)) destinos ,
( select * from transporte_dinamico .np_movimentacao
where cartao_data between '2017-10-01' and '2017-10-07' and cartao_hora between '0:00' and '23:59') movimentacoes
where st_Contains ( origens .contornoPg, movimentacoes.origem) and st_Contains ( destinos .contornoPg,
movimentacoes.destino)
order by 1 asc
```

Consulta B

```
select movimentacoes.cartao_sexo, movimentacoes.idade, movimentacoes.cartao_data ,
to_char (movimentacoes.cartao_datahora , 'DD/MM/YYYY, HH24h') as datahora_formatada,
movimentacoes.origem_lat, movimentacoes.origem_lng, movimentacoes. destino_lat , movimentacoes.destino_lng
from
( select id , ROI as contornoPg from transporte_dinamico . yp_ROIs where id in (2029)) origens ,
( select id , ROI as contornoPg from transporte_dinamico . yp_ROIs where id in (2013)) destinos ,
( select * from transporte_dinamico .np_movimentacao where cartao_data between '2017-10-01'
and '2017-10-07' and cartao_hora between '0:00' and '23:59' and upper(cartao_sexo) = 'F') movimentacoes
where st_Contains ( origens .contornoPg, movimentacoes.origem) and st_Contains ( destinos .contornoPg,
movimentacoes.destino)
order by 1 asc
```

Consulta C

```
select movimentacoes.cartao_sexo, movimentacoes.idade, movimentacoes.cartao_data ,
to_char (movimentacoes.cartao_datahora , 'DD/MM/YYYY, HH24h') as datahora_formatada,
movimentacoes.origem_lat, movimentacoes.origem_lng, movimentacoes. destino_lat , movimentacoes.destino_lng
from
( select id , ROI as contornoPg from transporte_dinamico . yp_ROIs where id in (2029,2062,2060,2073)) origens ,
( select id , ROI as contornoPg from transporte_dinamico . yp_ROIs where id in (2013,2028,2009,2083)) destinos ,
( select * from transporte_dinamico .np_movimentacao where cartao_data between '2017-10-01'
and '2017-10-07' and cartao_hora between '0:00' and '23:59') movimentacoes
where st_Contains ( origens .contornoPg, movimentacoes.origem) and st_Contains ( destinos .contornoPg,
movimentacoes.destino)
order by 1 asc
```

Consulta D

```
select movimentacoes.cartao_sexo, movimentacoes.idade, movimentacoes.cartao_data ,
to_char (movimentacoes.cartao_datahora , 'DD/MM/YYYY, HH24h') as datahora_formatada,
movimentacoes.origem_lat, movimentacoes.origem_lng, movimentacoes. destino_lat , movimentacoes.destino_lng
from
( select id , ROI as contornoPg from transporte_dinamico . yp_ROIs where id in (2029,2062,2060,2073)) origens ,
( select id , ROI as contornoPg from transporte_dinamico . yp_ROIs where id in (2013,2028,2009,2083)) destinos ,
( select * from transporte_dinamico .np_movimentacao where cartao_data between '2017-10-01'
and '2017-10-07' and cartao_hora between '0:00' and '23:59' and upper(cartao_sexo) = 'F') movimentacoes
where st_Contains ( origens .contornoPg, movimentacoes.origem) and st_Contains ( destinos .contornoPg,
movimentacoes.destino)
order by 1 asc
```

Consulta E

```
select movimentacoes.cartao_sexo, movimentacoes.idade, movimentacoes.cartao_data ,
to_char (movimentacoes.cartao_datahora , 'DD/MM/YYYY, HH24h') as datahora_formatada,
movimentacoes.origem_lat, movimentacoes.origem_lng, movimentacoes. destino_lat , movimentacoes.destino_lng
from
( select id , ROI as contornoPg from transporte_dinamico . yp_ROIs where id in (2051)) origens ,
( select id , ROI as contornoPg from transporte_dinamico . yp_ROIs where id in (2088)) destinos ,
( select * from transporte_dinamico .np_movimentacao where cartao_data between '2017-10-01'
and '2017-10-15' and cartao_hora between '8:00' and '20:00') movimentacoes
where st_Contains ( origens .contornoPg, movimentacoes.origem) and st_Contains ( destinos .contornoPg,
movimentacoes.destino)
order by 1 asc
```

Consulta F

```
select movimentacoes.cartao_sexo, movimentacoes.idade, movimentacoes.cartao_data ,
to_char (movimentacoes.cartao_datahora , 'DD/MM/YYYY, HH24h') as datahora_formatada,
movimentacoes.origem_lat, movimentacoes.origem_lng, movimentacoes. destino_lat , movimentacoes.destino_lng
from
( select id , ROI as contornoPg from transporte_dinamico . yp_ROIs where id in (2051)) origens ,
( select id , ROI as contornoPg from transporte_dinamico . yp_ROIs where id in (2088)) destinos ,
( select * from transporte_dinamico .np_movimentacao where cartao_data between '2017-10-01'
and '2017-10-15' and cartao_hora between '8:00' and '20:00' and upper(cartao_sexo) = 'F') movimentacoes
where st_Contains ( origens .contornoPg, movimentacoes.origem) and st_Contains ( destinos .contornoPg,
movimentacoes.destino)
order by 1 asc
```

Consulta G

```
select movimentacoes.cartao_sexo, movimentacoes.idade, movimentacoes.cartao_data ,
to_char (movimentacoes.cartao_datahora , 'DD/MM/YYYY, HH24h') as datahora_formatada,
movimentacoes.origem_lat, movimentacoes.origem_lng, movimentacoes. destino_lat , movimentacoes.destino_lng
from
( select id , ROI as contornoPg from transporte_dinamico . yp_ROIs where id in (2051)) origens ,
( select id , ROI as contornoPg from transporte_dinamico . yp_ROIs where id in (2088)) destinos ,
( select * from transporte_dinamico .np_movimentacao where cartao_data between '2017-10-01'
and '2017-10-15' and cartao_hora between '8:00' and '20:00' and idade >= 18 and idade < 65) movimentacoes
where st_Contains ( origens .contornoPg, movimentacoes.origem) and st_Contains ( destinos .contornoPg,
movimentacoes.destino)
order by 1 asc
```

Consulta H

```
select movimentacoes.cartao_sexo, movimentacoes.idade, movimentacoes.cartao_data ,
to_char (movimentacoes.cartao_datahora , 'DD/MM/YYYY, HH24h') as datahora_formatada,
movimentacoes.origem_lat, movimentacoes.origem_lng, movimentacoes. destino_lat , movimentacoes.destino_lng
from
( select id , ROI as contornoPg from transporte_dinamico . yp_ROIs where id in (2051)) origens ,
( select id , ROI as contornoPg from transporte_dinamico . yp_ROIs where id in (2088)) destinos ,
( select * from transporte_dinamico .np_movimentacao where cartao_data between '2017-10-01'
and '2017-10-15' and cartao_hora between '8:00' and '20:00' and idade >= 18 and idade < 65
and upper(cartao_sexo) = 'F') movimentacoes
where st_Contains ( origens .contornoPg, movimentacoes.origem) and st_Contains ( destinos .contornoPg,
movimentacoes.destino)
order by 1 asc
```

Consulta I

```
select movimentacoes.cartao_sexo, movimentacoes.idade, movimentacoes.cartao_data ,
to_char (movimentacoes.cartao_datahora , 'DD/MM/YYYY, HH24h') as datahora_formatada,
movimentacoes.origem_lat, movimentacoes.origem_lng, movimentacoes. destino_lat , movimentacoes.destino_lng
from
( select id , ROI as contornoPg from transporte_dinamico . yp_ROIs where id in (2054,2051)) origens ,
( select id , ROI as contornoPg from transporte_dinamico . yp_ROIs where id in (2088,2053)) destinos ,
( select * from transporte_dinamico .np_movimentacao where cartao_data between '2017-09-30' and '2017-10-30'
and cartao_hora between '0:00' and '23:59') movimentacoes
where st_Contains ( origens .contornoPg, movimentacoes.origem) and st_Contains ( destinos .contornoPg,
movimentacoes.destino)
order by 1 asc
```

Consulta J

```
select movimentacoes.cartao_sexo, movimentacoes.idade, movimentacoes.cartao_data ,
to_char (movimentacoes.cartao_datahora , 'DD/MM/YYYY, HH24h') as datahora_formatada,
movimentacoes.origem_lat, movimentacoes.origem_lng, movimentacoes. destino_lat , movimentacoes.destino_lng
from
( select id , ROI as contornoPg from transporte_dinamico . yp_ROIs where id in (2054,2051)) origens ,
( select id , ROI as contornoPg from transporte_dinamico . yp_ROIs where id in (2088,2053)) destinos ,
( select * from transporte_dinamico .np_movimentacao where cartao_data between '2017-09-30'
and '2017-10-30' and cartao_hora between '0:00' and '23:59' and upper(cartao_sexo) = 'F') movimentacoes
where st_Contains ( origens .contornoPg, movimentacoes.origem) and st_Contains ( destinos .contornoPg,
movimentacoes.destino)
order by 1 asc
```

Consulta K

```
select movimentacoes.cartao_sexo, movimentacoes.idade, movimentacoes.cartao_data ,
to_char (movimentacoes.cartao_datahora , 'DD/MM/YYYY, HH24h') as datahora_formatada,
movimentacoes.origem_lat, movimentacoes.origem_lng, movimentacoes. destino_lat , movimentacoes.destino_lng
from
( select id , ROI as contornoPg from transporte_dinamico . yp_ROIs where id in (2054,2051)) origens ,
( select id , ROI as contornoPg from transporte_dinamico . yp_ROIs where id in (2088,2053)) destinos ,
( select * from transporte_dinamico .np_movimentacao where cartao_data between '2017-09-30'
and '2017-10-30' and cartao_hora between '0:00' and '23:59' and idade >= 18 and idade < 65) movimentacoes
where st_Contains ( origens .contornoPg, movimentacoes.origem) and st_Contains ( destinos .contornoPg,
movimentacoes.destino)
order by 1 asc
```

Consulta L

```
select movimentacoes.cartao_sexo, movimentacoes.idade, movimentacoes.cartao_data ,
to_char (movimentacoes.cartao_datahora , 'DD/MM/YYYY, HH24h') as datahora_formatada,
movimentacoes.origem_lat, movimentacoes.origem_lng, movimentacoes. destino_lat , movimentacoes.destino_lng
from
( select id , ROI as contornoPg from transporte_dinamico . yp_ROIs where id in (2054,2051)) origens ,
( select id , ROI as contornoPg from transporte_dinamico . yp_ROIs where id in (2088,2053)) destinos ,
( select * from transporte_dinamico .np_movimentacao where cartao_data between '2017-09-30'
and '2017-10-30' and cartao_hora between '0:00' and '23:59' and upper(cartao_sexo) = 'F'
and idade >= 18 and idade < 65) movimentacoes
where st_Contains ( origens .contornoPg, movimentacoes.origem) and st_Contains ( destinos .contornoPg,
movimentacoes.destino)
order by 1 asc
```

APÊNDICE G - PROPÓSITO DE ALGUNS ARQUIVOS DO PROTÓTIPO

- **index.php:** é o arquivo inicial da aplicação. É responsável por dividir a interface de usuário verticalmente ao meio, estabelecendo à esquerda da tela a região do mapa e do menu, e à direita da tela a região dos gráficos
- **tela_inicial.php:** é o arquivo responsável pela construção do menu de busca da aplicação
- **bd_rois.php:** contém o *script* escrito em linguagem *PHP* responsável por montar sentença *SQL* responsável buscar na tabela transporte_dinamico.yp_rois os dados referentes aos contornos das Regiões de Interesse oferecidas pela aplicação para explorar os dados de OD do STP de Curitiba
- **bd_terminais.php:** contém o *script* escrito em linguagem *PHP* responsável por montar sentença *SQL* responsável buscar na tabela transporte_dinamico.yp_terminais os dados referentes aos terminais de embarque e desembarque do STP de Curitiba
- **bd_movimentacao.php:** contém o *script* escrito em linguagem *PHP* responsável por montar sentença *SQL* responsável buscar na tabela transporte_dinamico.np_movimentacao os dados referentes ao embarque e desembarque dos usuários do STP de Curitiba
- **bd_pontos_de_onibus.php:** contém o *script* escrito em linguagem *PHP* responsável por montar sentença *SQL* responsável buscar na tabela transporte_dinamico.yp_pontos_onibus os dados referentes aos pontos de ônibus do STP de Curitiba
- **index_main.php:** é arquivo mais importante da aplicação. Nele está contida toda a lógica de programação referente a criação dos gráficos, criação do mapa e também das camadas que podem ser adicionadas e removidas do mapa. As consultas quando disparadas são realizadas via chamada *JQuery.ajax*, a qual requisita os arquivos bd_XXXXXX.php para obter os dados armazenados na base.

APÊNDICE H - IMPLEMENTAÇÃO EM *R* PARA O *K-MEANS*

```
#carrega a lib para conexco no banco
library(RPostgreSQL)

#conectando no banco
conexao <- dbConnect(PostgreSQL(), dbname = "bigsea", host = "localhost",  port = 5435, user = "postread", password =
"PostRead")

#verifica se a tabela existe. Se sim, dropa a tabela
if(dbExistsTable(conexao, c('public', 'yp_movimentacao_com_clusters_v2'))){
  dbRemoveTable(conexao, c('public', 'yp_movimentacao_com_clusters_v2'))
}

#monta o SQL
SQL = "SELECT codlinha, nomelinha, codveiculo, numerocartao as cartao_numero, datautilizacao as cartao_datahora,
    to_date(to_char(datautilizacao, 'YYYY-MM-DD'), 'YYYY-MM-DD') as cartao_data,
    to_char(datautilizacao, 'HH24:MI:SS') as cartao_hora, data_nascimento as cartao_data_nascimento,
    sexo as cartao_sexo, origem_lat, origem_lng, origem_datahora, destino_lat, destino_lng, destino_datahora
    FROM public.yp_movimentacao"

#rodando o SQL
resultado = dbGetQuery(conexao, SQL)

#clusterizando origens com base na lat/lng dos embarques
k_05 <- kmeans(resultado[,10:11], 05, nstart = 10, iter.max = 100)
k_10 <- kmeans(resultado[,10:11], 10, nstart = 20, iter.max = 400)
k_20 <- kmeans(resultado[,10:11], 20, nstart = 40, iter.max = 1600)
k_30 <- kmeans(resultado[,10:11], 30, nstart = 60, iter.max = 3600)
k_40 <- kmeans(resultado[,10:11], 40, nstart = 80, iter.max = 6400)
k_50 <- kmeans(resultado[,10:11], 50, nstart = 100, iter.max = 12800)

#juntando tudo em um unico dataset
resultado_final <- cbind(resultado     , k_05['cluster'])
resultado_final <- cbind(resultado_final, k_10['cluster'])
resultado_final <- cbind(resultado_final, k_20['cluster'])
resultado_final <- cbind(resultado_final, k_30['cluster'])
resultado_final <- cbind(resultado_final, k_40['cluster'])
resultado_final <- cbind(resultado_final, k_50['cluster'])

#define os nomes para as colunas do dataset
nomes_colunas <- c("codlinha", "nomelinha", "codveiculo",
                   "cartao_numero", "cartao_datahora", "cartao_data", "cartao_hora",
                   "cartao_data_nascimento", "cartao_sexo",
                   "origem_lat", "origem_lng", "origem_datahora",
                   "destino_lat", "destino_lng", "destino_datahora",
                   "k_05", "k_10", "k_20", "k_30", "k_40", "k_50")

#renomeia as colunas
colnames(resultado_final) <- nomes_colunas

#define os tipos das colunas do dataset
tipos_colunas <- c("codlinha" = "text", "nomelinha" = "text", "codveiculo" = "text", "cartao_numero" = "text",
    "cartao_datahora" = "timestamp without time zone","cartao_data" = "date",
    "cartao_hora" = "time without time zone","cartao_data_nascimento" = "date",
    "cartao_sexo" = "text", "origem_lat" = "double precision", "origem_lng" = "double precision",
    "origem_datahora" = "timestamp without time zone", "destino_lat" = "double precision",
    "destino_lng" = "double precision", "destino_datahora" = "timestamp without time zone",
    "k_05" = "integer", "k_10" = "integer", "k_20" = "integer", "k_30" = "integer", "k_40" = "integer",
    "k_50" = "integer")

#adiciona dados na tabela do banco
dbWriteTable(conexao, c('public', 'yp_movimentacao_com_clusters_v2'), resultado_final, field.types = tipos_colunas,
row.names = FALSE)
```

APÊNDICE I - IMPLEMENTAÇÃO EM *R* PARA O *ELBOW METHOD*

```
#carrega a lib para conexao no banco
library("RPostgreSQL")

#conectando no banco
conexao <- dbConnect(PostgreSQL(), dbname = "BIGSEA", host = "localhost",
   port = 5434, user = "postread", password = "PostRead")

#consultando a tabela
dados_dia_03 = dbGetQuery(conexao, "SELECT latitude, longitude FROM transporte_dinamico.y_cartao_onibus
        WHERE cast(datautilizacao as DATE) = '2017-10-03'")

dados_dia_04 = dbGetQuery(conexao, "SELECT latitude, longitude FROM transporte_dinamico.y_cartao_onibus
        WHERE cast(datautilizacao as DATE) = '2017-10-04'")

dados_dia_05 = dbGetQuery(conexao, "SELECT latitude, longitude FROM transporte_dinamico.y_cartao_onibus
        WHERE cast(datautilizacao as DATE) = '2017-10-05'")

dados_dia_06 = dbGetQuery(conexao, "SELECT latitude, longitude FROM transporte_dinamico.y_cartao_onibus
        WHERE cast(datautilizacao as DATE) = '2017-10-06'")

dados_dia_07 = dbGetQuery(conexao, "SELECT latitude, longitude FROM transporte_dinamico.y_cartao_onibus
        WHERE cast(datautilizacao as DATE) = '2017-10-07'")

dados_dia_08 = dbGetQuery(conexao, "SELECT latitude, longitude FROM transporte_dinamico.y_cartao_onibus
        WHERE cast(datautilizacao as DATE) = '2017-10-08'")

dados_dia_09 = dbGetQuery(conexao, "SELECT latitude, longitude FROM transporte_dinamico.y_cartao_onibus
        WHERE cast(datautilizacao as DATE) = '2017-10-09'")

#remove os registros de NA dos datasets
dados_limpos_dia_03 = na.omit(dados_dia_03)
dados_limpos_dia_04 = na.omit(dados_dia_04)
dados_limpos_dia_05 = na.omit(dados_dia_05)
dados_limpos_dia_06 = na.omit(dados_dia_06)
dados_limpos_dia_07 = na.omit(dados_dia_07)

#definindo as funcoes para o Elbow de cada dia
tot_withinss_dia_03 <- function(k){ return (kmeans(dados_limpos_dia_03, k, nstart=1, iter.max = 60)tot.withinss)}
tot_withinss_dia_04 <- function(k){ return (kmeans(dados_limpos_dia_04, k, nstart=1, iter.max = 60)tot.withinss)}
tot_withinss_dia_05 <- function(k){ return (kmeans(dados_limpos_dia_05, k, nstart=1, iter.max = 60)tot.withinss)}
tot_withinss_dia_06 <- function(k){ return (kmeans(dados_limpos_dia_06, k, nstart=1, iter.max = 60)tot.withinss)}
tot_withinss_dia_07 <- function(k){ return (kmeans(dados_limpos_dia_07, k, nstart=1, iter.max = 60)tot.withinss)}

#chama as funcoes criadas e obtem a lista de valores para plotagem do Elbow
Elbow_dia_03 <- sapply(1:40, tot_withinss_dia_03)
Elbow_dia_04 <- sapply(1:40, tot_withinss_dia_04)
Elbow_dia_05 <- sapply(1:40, tot_withinss_dia_05)
Elbow_dia_06 <- sapply(1:40, tot_withinss_dia_06)
Elbow_dia_07 <- sapply(1:40, tot_withinss_dia_07)

#define o nome e local onde o grafico sera gerado
jpeg(file= "e:\\Elbow_multidias_dados_brutos.jpeg", width=4000, height=3000, units = 'px', res = 300 )

#plota o grafico
plot(Elbow_dia_03, type="o", col = "red", xlab="Numero de clusters",
   ylab="Total within-clusters sum of squares", pch = 1, cex = 2)
lines(Elbow_dia_04, type = "o", col = "blue", pch = 2, cex = 2)
lines(Elbow_dia_05, type = "o", col = "green", pch = 3, cex = 2)
lines(Elbow_dia_06, type = "o", col = "yellow", pch = 4, cex = 2)
lines(Elbow_dia_07, type = "o", col = "cyan", pch = 5, cex = 2)
legend("topright", legend = c('Dia 3', 'Dia 4', 'Dia 5', 'Dia 6', 'Dia 7'),
      col = c('red', 'blue', 'green', 'yellow', 'cyan'), pch = c(1,2,3,4,5), cex = 3)

#desabilita a sa?da no v?deo para que o grafico seja gerado em arquivo
```

APÊNDICE J - FORMULÁRIO DE PESQUISA PÓS-PROTÓTIPO

Questionário

As respostas deste questionário são anônimas e serão utilizadas para a auxiliar o desenvolvimento de um projeto de mestrado do PPGCA da UTFPR. Desde já, agradecemos pela colaboração.

1. **Você foi capaz de realizar uma pesquisa de Origem-Destino na nossa ferramenta?**
   Marcar apenas uma oval.

   ◯ Sim ◯ Não

2. **Avalie a facilidade de uso da ferramenta:**
   Marcar apenas uma oval.

	1	2	3	4	5	
Péssimo	◯	◯	◯	◯	◯	Ótimo

3. **Avalie as opções de filtros oferecidos pela ferramenta:**
   Marcar apenas uma oval.

	1	2	3	4	5	
Péssimo	◯	◯	◯	◯	◯	Ótimo

4. **Avalie a forma como os resultados são apresentados pela ferramenta:**
   Marcar apenas uma oval.

	1	2	3	4	5	
Péssimo	◯	◯	◯	◯	◯	Ótimo

5. **O que poderia ser melhorado na ferramenta?**

_____

_____

_____

APÊNDICE K - ROTEIRO DO TESTE DE USABILIDADE

Nível 1

1) Quem mais viajou do bairro CIC até o bairro Centro desde 01/10/2017 a 01/10/2017, das 00:00 as 23:59: homens ou mulheres?

Nível 2

3) Qual a faixa etária que mais viaja da regional Matriz a regional CIC desde 10/10/2017 a 20/10/2017, das 19:00 as 23:59?

Nível 3

5) Obtenha uma visão clusterizada dos embarques dos deslocamentos do bairro CIC para o bairro Batel desde 01/10/2017 a 25/10/2017, das 06:00 as 20:00