



Revista de Administração de Empresas

ISSN: 0034-7590

ISSN: 2178-938X

Fundação Getúlio Vargas, Escola de Administração de
Empresas de S.Paulo

INSARDI, ANDRÉ; LORENZO, RODOLFO OLIVEIRA
MEASURING ACCESSIBILITY: A BIG DATA PERSPECTIVE ON UBER SERVICE WAITING TIMES
Revista de Administração de Empresas, vol. 59, no. 6, 2019, November-December, pp. 402-414
Fundação Getúlio Vargas, Escola de Administração de Empresas de S.Paulo

DOI: 10.1590/S0034-759020190606

Available in: <http://www.redalyc.org/articulo.oa?id=155163302006>

- How to cite
- Complete issue
- More information about this article
- Journal's webpage in redalyc.org

UABM redalyc.org

Scientific Information System Redalyc
Network of Scientific Journals from Latin America and the Caribbean, Spain and
Portugal

Project academic non-profit, developed under the open access initiative

FÓRUM

Submetido 19.10.2018. Aprovado 19.07.2019

Avaliado pelo sistema *double blind review*. Editores Científicos Convidados: Eduardo de Rezende Francisco, José Luiz Kugler, Soong Moon Kang, Ricardo Silva e Peter Alexander Whigham

Versão traduzida

DOI: <http://dx.doi.org/10.1590/S0034-759020190606>

MEDINDO A ACESSIBILIDADE: UMA PERSPECTIVA DE *BIG DATA* SOBRE OS TEMPOS DE ESPERA DO SERVIÇO DA UBER

Measuring accessibility: a Big Data perspective on uber service waiting times

Medición de accesibilidad: Una perspectiva de Big Data sobre los tiempos de espera del servicio de la Uber

RESUMO

O presente artigo busca relacionar informações sobre o tempo de espera de serviços de aluguel de carro, especificamente Uber, com variáveis socioeconômicas da cidade de São Paulo, com a intenção de explorar a possibilidade de uso dessas medidas como um proxy de acessibilidade. Foi montada uma base com a média dos dados de tempo de espera do serviço por distrito, que foi agregada a um conjunto de variáveis socioeconômicas e de infraestrutura de transporte. A partir dessa base, foram elaborados modelos de regressão linear múltipla (RLM), e, utilizando o método stepwise, foram selecionadas as variáveis mais significativas do modelo. Foi verificado o padrão espacial das variáveis por meio do teste I de Moran, que motivou a elaboração de um modelo espacial autorregressivo (SAR). Os resultados indicam que variáveis físicas são importantes para essa relação, como área e densidade populacional, mas a quilometragem de linhas de ônibus no distrito a taxa de residentes não brancos, além do componente espacial, indica uma possível relação com acessibilidade.

PALAVRAS-CHAVE | Acessibilidade, *Big Data*, Uber, estatística espacial, disparidade urbana.

ABSTRACT

This study aims to relate information about the waiting times of ride-sourcing services, with specific reference to Uber, using socioeconomic variables from São Paulo, Brazil. The intention is to explore the possibility of using this measure as an accessibility proxy. A database was created with the mean waiting time data per district, which was aggregated to a set of socioeconomic and transport infrastructure variables. From this database, a multiple linear regression model was built. In addition, the stepwise method selected the most significant variables. Moran's I test confirmed the spatial distribution pattern of the measures, motivating the use of a spatial autoregressive model. The results indicate that physical variables, such as area and population density, are important to explain this relation. However, the mileage of district bus lines and the non-white resident rate were also significant. Besides, the spatial component indicates a possible relation to accessibility.

KEYWORDS | Accessibility, *Big Data*, Uber, space statistic, urban disparity.

RESUMEN

El presente artículo busca relacionar informaciones sobre el tiempo de espera de servicios de alquiler de coches, específicamente Uber, con variables socioeconómicas de la ciudad de São Paulo con la intención de explorar la posibilidad de utilizar esas medidas como un proxy de accesibilidad. Se ha montado una base con la media de los datos de tiempo de espera del servicio por distrito, que se ha agregado a un conjunto de variables socioeconómicas y de infraestructura de transporte. A partir de esta base se elaboraron modelos de regresión lineal MLR, y utilizando el método stepwise se seleccionaron las variables más significativas del modelo. Se verificó el patrón espacial de las variables a través de la prueba I de Moran, que motivó la elaboración de un modelo espacial autoregresivo (SAR). Los resultados indican que las variables físicas son importantes para esa relación, como el área y la densidad de población, pero el kilometraje de líneas de autobús en el distrito, la tasa de residentes no blancos, además del componente espacial, indica una posible relación con accesibilidad.

PALABRAS CLAVE | Accesibilidad, *Big Data*, Uber, estadística espacial, disparidad urbana.

ANDRÉ INSARDI¹

andre.insardi@espm.br

ORCID: 0000-0003-3782-3505

RODOLFO OLIVEIRA LORENZO²

rodolfo@uol.com.br

ORCID: 0000-0003-4847-9201

¹ Escola Superior de Propaganda e Marketing, São Paulo, SP, Brasil

² Fundação Getúlio Vargas, Escola de Administração de Empresas de São Paulo, São Paulo, São Paulo, Brasil

INTRODUÇÃO

A introdução de empresas “ride-sourcing” (empresas de terceirização de viagens) no mercado privado de mobilidade urbana mudou consideravelmente os hábitos de muitos moradores das cidades, inclusive dos seus utilizadores tradicionais, os proprietários de automóveis particulares e passageiros.

O aumento da concorrência num mercado que tradicionalmente é altamente regulamentado levou a conflitos entre os novos serviços, os antigos fornecedores desse mercado, os táxis e as autoridades locais em muitas cidades. As dificuldades enfrentadas pelas autoridades para enquadrar esses novos serviços no quadro jurídico existente são claramente ilustradas no caso de São Francisco (Flores & Rayle, 2017).

Nesse contexto, empresas de terceirização de viagens tentaram construir uma imagem de serem amigas do ambiente, vendendo-se como serviços de compartilhamento de viagens (Flores & Rayle, 2017). Além disso, elas reivindicaram a capacidade de reduzir o número de veículos da rua (e suas emissões) e de oferecer serviços melhores e mais baratos a áreas anteriormente negligenciadas pelos táxis.

Nesse contexto, vários estudos procuraram esclarecer os efeitos das transformações resultantes. Jin, Kong, Wu e Sui (2018) revisaram estudos recentes sobre os efeitos dos serviços de terceirização de viagens. Entre os assuntos discutidos, estão a eficiência econômica e a equidade social desses serviços. Em termos gerais, parece haver evidência do aumento dos serviços de transporte privado em áreas periféricas e de baixa renda anteriormente negligenciadas pelos táxis. Apesar disso, o perfil médio do cliente do serviço é mais jovem, mais rico e mais educado do que o perfil médio da população; o artigo aponta para uma possível exclusão do serviço ao estilo de “divisão digital”, tanto em termos de geração como de renda (Jin et al., 2018).

No entanto, outro aspecto essencial dos serviços de terceirização de viagens é a sua capacidade de gerar *Big Data* de alta qualidade que pode ser usado para avaliar várias questões relativas ao serviço. O acesso a esses dados permite uma análise rica e detalhada e pode revelar ainda mais sobre a interação entre outros modos de transporte e terceirização de viagens (Jin et al., 2018).

Em especial, a possibilidade de analisar a acessibilidade urbana a partir dessa fonte de dados é interessante. Como componente essencial da acessibilidade, a distribuição da rede de transportes (Páez, Scott, & Morency, 2012) é um tema de pesquisa recorrente entre geógrafos, urbanistas e cientistas sociais. Recentemente, os métodos e a abordagem desse campo de estudos estão sendo transformados. Isso é resultado da influência de, entre outras coisas, *Big Data* e abertura de diálogos com

outras disciplinas (Schwanen, 2016). Ao mesmo tempo, o campo continua a tornar-se relevante: enquanto o atrito da distância for uma realidade, a acessibilidade continuará sendo um conceito útil para descrever a experiência urbana (Páez et al., 2012).

Considerando principalmente abordagens de *Big Data*, Letouzé e Jütting (2015) afirmam que os órgãos oficiais responsáveis pela produção de estatísticas e indicadores oficiais, incluindo instituições acadêmicas, devem estar cientes das evoluções do *Big Data*. Isso é necessário tanto para lucrar com novas ferramentas e abordagens, o que vem junto de rigor científico da validação e análise, como para encarar o mundo do *Big Data* como uma fonte inestimável de dados para o avanço da pesquisa científica. O potencial dos instrumentos de *Big Data* também é de fundamental importância para os gestores públicos e os decisores políticos (Kim, Trimi, & Chung, 2014). A utilização de diferentes ferramentas de *Big Data* numa rede de transportes multimodais pode ajudar a regular melhor a oferta de transporte privado e a fornecer melhor transporte público de acordo com as necessidades dos utilizadores (Kim et al., 2014; Lessa, Lobo, & Cardoso, 2019).

Nesse contexto, o crescimento das empresas tecnológicas globais de mobilidade, tais como Uber, Cabify e Lyft, tornou-as atores consideráveis nessa área. Por exemplo, a Uber realiza uma média diária de 15 milhões de corridas em todo o mundo. No Brasil, a empresa oferece seu serviço em 100 cidades, com uma rede de 500 mil motoristas e mais de 20 milhões de usuários.

A Uber desenvolve, comercializa e opera o aplicativo para *smartphones* que permite aos consumidores solicitar corridas de motoristas parceiros. No processo de solicitação de corridas, a ferramenta Uber fornece estimativas de tempo de espera e custo de viagem no aplicativo do usuário e no ambiente “web” por meio de interfaces de programação de aplicações (APIs) públicas.

Essas APIs geram estimativas por meio da análise do histórico de corridas na região do usuário e da curva de oferta e demanda dos carros da Uber (Cohen et al., 2016), conforme pode ser verificado na documentação disponível (<https://developer.uber.com>).

Wang & Mu (2018) e Hughes & Mackenzie (2016) propõem o uso dessas estimativas como uma possível medida de acessibilidade. As interferências desses novos serviços no ambiente de transporte e o fácil acesso às ferramentas que a Uber fornece já motivaram a realização de alguns estudos (Hall & Krueger, 2016; Hughes & MacKenzie, 2016; Wang & Mu, 2018; Zhou, Wang, & Li, 2017). Isso abre oportunidade para a exploração empírica da frota da Uber à luz da teoria da acessibilidade.

O objetivo deste estudo é utilizar as ferramentas de *Big Data* desenvolvidas pela Uber, uma das maiores provedoras de terceirização de viagens de São Paulo, para gerar dados para a

realização de um estudo exploratório possivelmente mensurando acessibilidade.

Como o algoritmo de preços da Uber segue o equilíbrio entre oferta e procura, a fim de influenciar o comportamento do motorista (Hall, Horton, & Knoepfle, 2019), inferimos que os tempos de espera para corridas na Uber podem refletir desequilíbrios regionais na oferta de carros. Seguindo essa linha de pensamento, exploramos a relação dos tempos de espera com outras variáveis associadas à acessibilidade em literatura similar (Hughes & MacKenzie, 2016; Lessa et al., 2019; Wang & Mu, 2018). Em especial, variáveis relacionadas a mobilidade e fatores socioeconômicos.

Os resultados encontrados são divergentes dos encontrados em estudos anteriores (Hughes & MacKenzie, 2016; Wang & Mu, 2018) em relação à relevância das variáveis no contexto de São Paulo. Tendo em mente o alerta feito por Schwanen (2016) sobre a generalização de conclusões advindas do uso de *Big Data* em diferentes contextos locais, estudos adicionais podem comparar diferentes cidades para elucidar como o contexto local pode ser mais bem considerado em abordagens semelhantes. Além disso, uma validação adicional dessa fonte de dados pode construir uma nova ferramenta capaz de facilitar a tomada de decisão dos planejadores de transportes e gestores de políticas públicas nas cidades (Kim et al., 2014; Letouzé & Jütting, 2015).

REVISÃO DA LITERATURA E QUESTÃO DE PESQUISA

Esta seção começa com a conceitualização de *Big Data* e seu posicionamento no mercado de tecnologia atual e a conceitualização da ferramenta de estimativa da Uber à luz do conceito de *Big Data*. Seguem-se uma revisão do conceito de acessibilidade, suas disposições normativas e positivistas e seu contexto gerencial.

Big Data

Uma definição clássica do movimento de *Big Data* leva em consideração as características dos dados produzidos nos ambientes virtuais com presença massiva do usuário: Volume, Variedade e Velocidade, os três Vs (McAfee & Brynjolfsson, 2012).

De acordo com essa definição, os dados gerados pelas novas formas de usar tecnologia e aplicativos geram bases de dados relativamente grandes, chegando à escala de Petabytes ou Exabytes. As diversas formas de geração de dados (como postagens com fotos, comentários em redes sociais, reações a comentários de outros, vídeos e áudios etc.) são responsáveis

pela geração de bases de dados heterogêneas, em contraste com as bases de dados estruturadas. A geração dessas bases de dados é quase instantânea, exigindo processamento em tempo real em alguns casos. Esses aspectos subsidiam as técnicas capazes de lidar com os dados neste período valioso.

Essa não é a única abordagem conceitual para *Big Data*. Uma visão mais sociológica tenta descrever o movimento com três “Cs”: “Crumbs”, Capacidades e Comunidades (Letouzé & Jütting, 2015).

“Crumbs”, ou migalhas, é uma referência à natureza dos dados coletados em relação ao comportamento dos usuários nessas novas aplicações. Esses usuários deixam vestígios de suas atividades enquanto interagem, e esses traços, ou migalhas, constituem os bancos de dados a serem analisados em *Big Data*. Capacidades são as técnicas, tanto estatísticas como de programação, utilizadas para manipular esses dados e extrair informações. O terceiro conceito, Comunidades, refere-se aos padrões de comportamento dos produtores de ambientes de *Big Data* dentro de comunidades específicas cujos membros compartilham ideias com linguagem específica e métodos de validação comuns. Essas comunidades podem ser estabelecidas em ambientes abertos e colaborativos, como comunidades open source, ou em grupos mais restritivos, como grupos de tecnologia em grandes corporações com acesso a grandes bancos de dados (Letouzé & Jütting, 2015).

A API da Uber no contexto de *Big Data*

A ferramenta de estimativa da Uber usada neste artigo é, por ambas as definições, uma ferramenta de *Big Data*. Em relação aos três Vs, o enorme volume de corridas faz com que os dados estejam constantemente alimentando o algoritmo da ferramenta com interpretações de alta velocidade de dados espaciais não estruturados (Cohen et al., 2016).

Ao mesmo tempo, os dados que alimentam o algoritmo são traços de atividades dos motoristas e usuários, ou migalhas. As ferramentas que processam os dados recebidos on-line a essa alta taxa também são ferramentas incluídas no termo Capacidades, desenvolvido por ambientes de *Big Data*. Além disso, a comunidade de operadores no sistema de startups de transporte, particularmente Uber, se encaixa no conceito de Comunidades.

Além disso, a descrição de Pääkkönen e Pakkala (2015) da arquitetura de software de *Big Data* se adapta muito bem à arquitetura e tecnologia da API da Uber, conforme bem descrito em sua documentação (<https://developer.uber.com>).

Acessibilidade

Discutir o acesso dos indivíduos à mobilidade urbana na cidade é um tema importante. Há uma série de abordagens diferentes usando desde questionários estatísticos tradicionais (Metrô, 2008) até simulações computacionais de comportamento de mobilidade urbana em determinado território urbano (Krajewicz, Erdmann, Behrisch, & Bieker, 2012), incluindo abordagens semelhantes à atual, usando dados de mídia social para traçar o comportamento de mobilidade (Noulas, Scellato, Lambiotte, Pontil, & Mascolo, 2012).

Em grandes centros urbanos, comparáveis em tamanho e importância regional a São Paulo, conhecer a mobilidade das pessoas é essencial para compreender a dinâmica urbana, pois o tamanho da cidade alimenta sua capacidade de atrair mais pessoas (Aranha, 2005). Isso tem implicações para soluções de mobilidade que têm como objetivo atender a demanda por locomoção. Um dos principais fatores para medir a capacidade de movimentação dos indivíduos dentro da cidade é o tempo usual das corridas deles.

Em São Paulo, essa variável reflete contextos socioeconômicos, indicando que o acesso à mobilidade se distribui de modo diferente em termos de espaço e contextos sociais (Morandi et al., 2016). No entanto, a mobilidade como conceito de capacidade de movimentação na cidade é limitada no sentido de não considerar as condições de locomoção.

Um conceito mais abrangente nesse sentido é o de acessibilidade (Litman, 2003; Páez et al., 2012; Stelder, 2016). É possível definir acessibilidade como o potencial de acesso a oportunidades distribuídas espacialmente (Wang & Mu, 2018). Hansen (1959) define acessibilidade como a maneira como as pessoas interagem com lugares. Para ilustrar, a quantidade de estradas medida em quilômetros em uma determinada região é uma medida comumente utilizada. Da mesma forma, na geografia, a acessibilidade é a medida de como uma pessoa participa de uma determinada atividade (Kwan, 1998; Weibull, 1980).

A acessibilidade reflete o desenvolvimento espacial, que consiste em redes de transporte e distribuição de oportunidades, materializadas em usos e ocupações do solo. A partir desse conceito, acessibilidade pode ser interpretada como uma medida temporal (tempo de acesso) (Lessa et al., 2019; Páez et al., 2012). Como exemplo prático, é possível considerar o tempo de viagem até o trabalho como uma medida comparativa para compreender o equilíbrio na ocupação profissional e as disparidades raciais, econômicas e de gênero contidas nas distribuições da área urbana (Preston & McLafferty, 1999; Tribby & Zandbergen, 2012).

Geógrafos e cientistas sociais têm analisado criticamente a economia e a desigualdade no transporte, e suas correlações

com as desigualdades socioeconômicas. Schwanen (2016) argumenta que a distribuição de transportes intensifica a polarização socioespacial sob a dinâmica do capitalismo, porque a infraestrutura de transporte é um ativo que atrai capital e investimentos, trazendo oportunidades de emprego, mais eficiência e concorrência.

Ao discutirem sobre acessibilidade, Páez et al. (2012) definem duas abordagens epistêmicas para estudar o conceito: a primeira é uma abordagem normativa definida em termos de quais parâmetros de acessibilidade devem ser considerados razoáveis ou, em outras palavras, o quanto é razoável para uma pessoa realizar uma viagem. A segunda abordagem é o positivismo, e é definida em termos de parâmetros de acessibilidade observados ou o quanto as pessoas realizam viagens. A abordagem normativa analisa as expectativas de viagem, enquanto o positivismo se baseia na experiência de viagem real. Nosso estudo levará em consideração a abordagem positivista, já que os dados se referem aos tempos reais de espera da Uber em São Paulo.

Estudos como os de Hughes e MacKenzie (2016), Lessa et al. (2019) e Wang e Mu (2018) também utilizam a abordagem positivista. De uma forma mais tradicional, Lessa et al. (2019) investigaram as relações entre os tempos de corrida coletados em questionários de origem-destino e dados de trânsito e a distribuição da infraestrutura da rede de transportes públicos. A abordagem de Hughes e MacKenzie (2016) para acessibilidade usa tempos de espera por serviços Uber e dados socioeconômicos para explorar as correlações espaciais entre a dispersão do serviço da Uber e as disparidades socioeconômicas de Seattle. Esse dado pode ser mensurado por variáveis como densidade populacional, renda média per capita e percentual de não brancos. Wang e Mu (2018) realizaram um estudo transversal, criando um modelo de regressão espacial que testa a relação do tempo de espera pelo Uber com variáveis socioeconômicas e de infraestrutura de transporte da cidade de Atlanta.

Como Hughes e MacKenzie (2016) e Wang e Mu (2018) apontam, esses estudos preliminares estão sujeitos a influências econômicas e culturais do local pesquisado. Além disso, Letouzé e Jütting (2015), Schwanen (2016) e Kwan (2016) nos alertam sobre a possível existência de viés em bancos de dados construídos a partir de ferramentas de *Big Data*. Isso é extremamente relevante para a discussão da acessibilidade e uso de bancos de dados provenientes de *Big Data* na investigação do tempo de espera pelo Uber como proxy para acessibilidade em São Paulo, pois o contexto econômico e cultural em questão é relevante para os resultados da análise.

Inspirado nesses estudos, este artigo discute as seguintes hipóteses no contexto de São Paulo:

H_1 : O tempo de espera estimado ao solicitar uma viagem na plataforma da Uber pode ser usado como proxy para acessibilidade.

H_2 : A distribuição do tempo de espera pelo Uber está associada à polarização dos indicadores socioeconômicos.

METODOLOGIA

Esta seção começa com uma descrição dos dados coletados a respeito dos tempos estimados de espera pelo Uber e dados socioeconômicos seguidos de uma análise descritiva para definir os recortes de dados para análise final (Wang & Mu, 2018). Em seguida, explicamos o processo de construção dos modelos de regressão linear múltipla (MLR) com método “stepwise”. Como as variáveis significativas selecionadas por esse método foram consideradas pelo teste I de Moran como tendo alta dependência espacial, calculamos um modelo espacial autorregressivo final (SAR), conforme sugerido por Wang e Mu (2018).

Dados e variáveis

Para verificar as hipóteses propostas acima, foram coletados dois conjuntos de dados. O primeiro são os dados dos tempos de espera pelo Uber e o segundo, dados socioeconômicos espacializados de São Paulo.

Dados Uber

Os dados foram coletados a partir das APIs presentes no portal de desenvolvedores da Uber, acessíveis a partir do *site* da empresa (<https://developer.uber.com>), os quais abrangem os tempos de espera estimados de todos os produtos Uber durante agosto de 2018.

A cidade de São Paulo está dividida em 96 distritos, uma divisão territorial e administrativa que proporciona à administração local um certo grau de autonomia (Francisco, 2010). Seguindo a metodologia proposta por Wang e Mu (2018), os distritos foram utilizados como unidade espacial de análise. Para garantir que cada distrito tivesse pelo menos um ponto amostral aleatório, o software desenvolvido para a coleta de dados seguiu a lógica abaixo:

- Seccionar a cidade em quadrados de 1 km², totalizando 1.720 quadrados;
- amostrar aleatoriamente um ponto em cada quadrado;
- as coordenadas amostradas são usadas para consultar a API de tempos de espera pelo Uber;
- se o retorno for bem-sucedido, o resultado é armazenado;

- se um erro for sinalizado, mais três tentativas com a mesma coordenada são feitas antes de a tentativa ser registrada como um nulo.

Esse processo foi repetido a cada 30 minutos durante agosto de 2018. Mais de 2.528.400 chamadas para a API foram armazenadas, sendo pouco mais de 2.240.000 válidas.

Para processar os dados coletados, um programa foi desenvolvido em C# seguindo os conceitos de computação em nuvem para acessar as APIs da Uber. Esse programa foi alocado no serviço de computação em nuvem do Azure fornecido pela Microsoft. Sua arquitetura é apresentada na Figura 1.

Dados socioeconômicos

A escolha dos indicadores socioeconômicos foi baseada em estudos semelhantes (Hughes & MacKenzie, 2016; Wang & Mu, 2018). As unidades de análise escolhidas foram os distritos de São Paulo. Foram encontradas variáveis comparáveis para densidade populacional, densidade de emprego, taxa de minoria, renda média per capita, taxa de motorização (para carros e motocicletas), infraestrutura de transporte público e tempo médio de viagem para o trabalho. Os dados socioeconômicos em nível distrital foram coletados a partir do censo demográfico do Instituto Brasileiro de Geografia e Estatística (IBGE, 2010), do Ministério Nacional do Trabalho e Emprego (2016, Relatório Anual de Informação Social – RAIS), do portal de indicadores dos municípios paulistas (IMP) do Seade e do portal municipal de informação georreferenciada de São Paulo (<http://geosampa.prefeitura.sp.gov.br>). Esses dados encontram-se resumidos no Quadro 1.

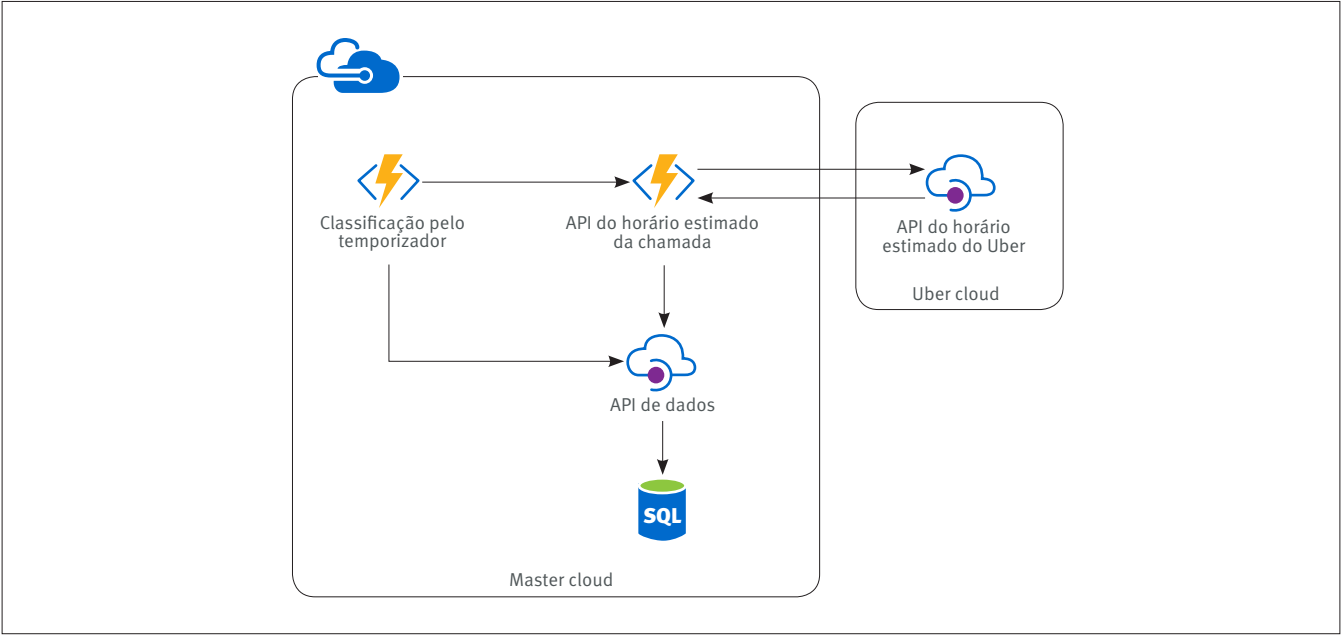
O censo demográfico decenal é um dos produtos estatísticos mais importantes do Brasil. O censo de 2010 apresenta dois conjuntos de dados: o universal, que idealmente compreende todos os brasileiros, e a amostra, na qual é feito um levantamento mais detalhado dos entrevistados (uma fração estatística das populações). A fração amostral varia entre municípios: em São Paulo, cerca de 5% dos domicílios foram incluídos na amostra. Em todo o Brasil, foram selecionados 10,7%, ou seja, 6.192.332 domicílios (IBGE, 2010).

A RAIS é a ferramenta do Ministério do Trabalho e Emprego para a gestão das relações de trabalho brasileiras. Os dados são compilados a partir de declarações feitas por empresas sobre a situação de trabalho entre elas e seus funcionários. As declarações são obrigatórias para uma série de empresas nomeadas por lei (Ministério do Trabalho, 2016).

A Fundação Seade é uma instituição estatística reconhecida nacionalmente. A fundação é conhecida por sua capacidade

técnica. Também é responsável, já que faz parte do Sistema Nacional de Estatística para produção de dados, pela agregação de dados existentes no interesse do Estado de São Paulo e de seus municípios (Francisco, 2010).

Figura 1. Diagrama da aplicação de coleta de dados da Uber



Quadro 1. Dados socioeconômicos de São Paulo

| Dados distritais | Fonte | Data |
|---|--|-----------|
| Área (km²) | Dados do IBGE. Coletados no portal IMP – Fundação Seade | 2009 |
| População | Dados originais do censo demográfico do IBGE reajustados anualmente pela Fundação Seade. Coletado no portal IMP – Fundação SEADE | 2010/2018 |
| Densidade da população | Calculada a partir de dados da área e população | 2010/2018 |
| Domicílios particulares permanentes | Dados originários do censo demográfico do IBGE reajustados anualmente pela Fundação Seade. Coletado no portal IMP – Fundação Seade | 2010/2018 |
| Rendimento <i>per capita</i> – censo demográfico (em reais atuais) | Dados originários do censo demográfico do IBGE. Coletado no portal IMP – Fundação Seade | 2010 |
| Empregos (comércio, serviços, indústria da transformação, construção civil) | Portal municipal de São Paulo "Infocidade". Fonte de dados original: Ministério do Trabalho e Emprego - RAIS | 2010/2016 |
| Empregadores (comércio, serviços, indústria da transformação, construção civil) | Portal municipal de São Paulo "Infocidade". Fonte de dados original: Ministério do Trabalho e Emprego - RAIS | 2010/2016 |
| % de não brancos (preto, pardos, indígenas) | Censo demográfico 2010 do IBGE | 2010 |
| Tempos de viagem | Amostra do censo demográfico do IBGE. Proporção de pessoas ponderadas em cada classe de tempo de viagem descrita no questionário, por distrito | 2010 |
| Taxa de motorização doméstica | Amostra do censo demográfico do IBGE. Taxa de motorização por distrito (carros e motocicletas) dos domicílios ponderados da amostra | 2010 |
| Paradas de ônibus | Portal Geosampa | 2018 |
| Extensão das linhas de ônibus (km) | Portal Geosampa | 2018 |
| Estações de metrô | Portal Geosampa | 2018 |

A plataforma on-line municipal de São Paulo Geosampa fornece uma série de dados georreferenciados sobre uma série de questões, incluindo a distribuição de infraestrutura de transporte.

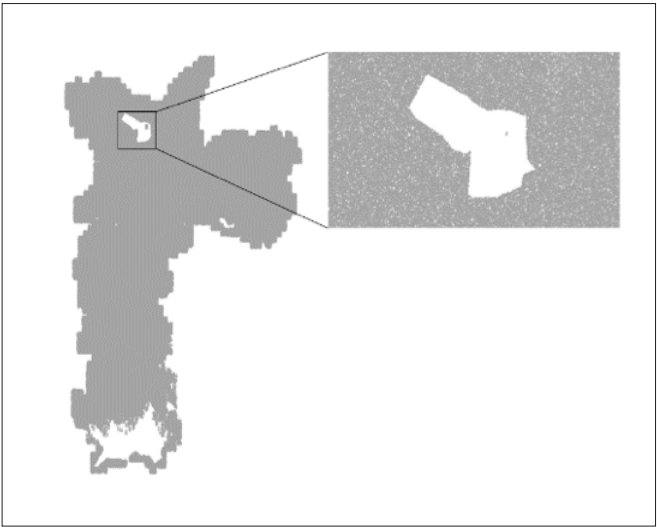
Todos os dados foram pesquisados ou agregados no nível do distrito (Francisco, 2010; Hughes & MacKenzie, 2016; Lessa et al., 2019; Wang & Mu, 2018). A Amostra Censo de 2010 foi utilizada para contabilizar o percentual da população de cada distrito que está em diferentes classes de tempo de viagem para trabalho/escola, e também o percentual de domicílios em cada distrito que possuem carro e motocicleta.

A distribuição de dados socioeconômicos segue um padrão claro de periferia central com pequenas variações, e está associada à distribuição de infraestrutura e políticas públicas (Francisco, 2010; Torres, Marques, Ferreira, & Bitar, 2003). As relações desse padrão com a distribuição dos tempos de espera pelo Uber podem lançar alguma luz sobre a relação de acessibilidade e composição socioespacial, como Wang e Mu (2018) investigaram.

Análise descritiva de dados

A exploração das 2.528.400 chamadas da API mostrou dois padrões: em primeiro lugar, a falta geral de estimativas para alguns dos produtos da Uber e, em segundo lugar, a ausência completa de chamadas em algumas regiões da cidade, como é mostrado na Figura 2. É possível observar a ausência no extremo sul da cidade e em uma faixa ao norte, além de alguns pontos espalhados pela cidade.

Figura 2. Pontos de coordenada para estimativas de tempo de espera pelo Uber



As causas identificadas foram a presença de uma área de mangue no sul sem acesso rodoviário. A faixa norte e os outros pontos sobre a cidade são regiões que a Uber classifica como zonas de risco, onde não presta serviços.

A cobertura de cada serviço e a sua disponibilidade podem ser vistas no Tabela 1. O UberX, serviço mais popular que a Uber oferece, tem a maior cobertura, com quase 100% das respostas estimadas de tempo registradas. Consequentemente, é o serviço que melhor reflete a distribuição espacial da frota da Uber. Portanto, os tempos de espera pelo UberX foram usados para medir acessibilidade.

Tabela 1. Retornos de chamadas de estimativas de tempo de espera pelo Uber, por produto

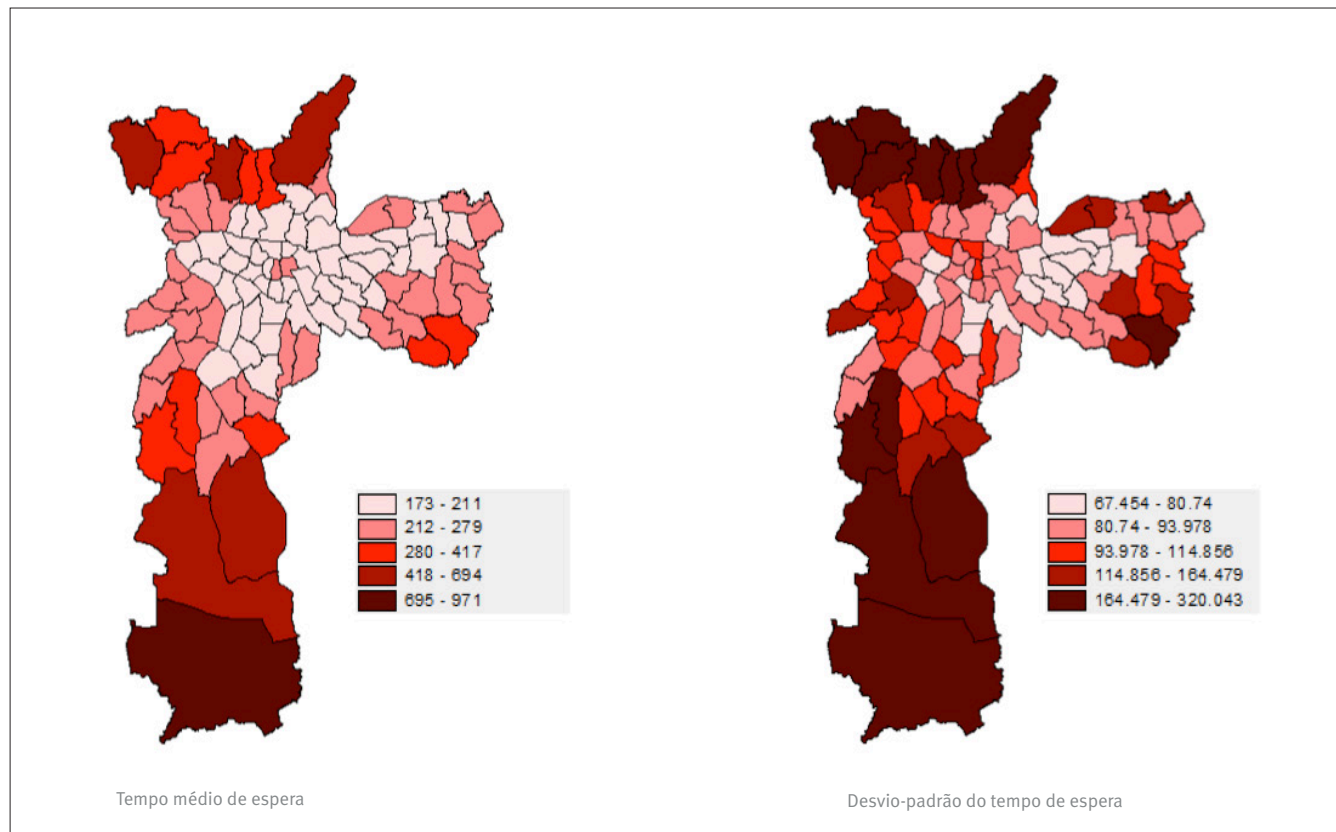
| Serviço | Número de chamadas bem-sucedidas | Média do tempo de espera estimado (segundos) | Desvio-padrão do tempo de espera estimado (segundos) |
|-----------|----------------------------------|--|--|
| Bag | 1.327.710 | 400 | 187,2 |
| Bike Rack | 58.758 | 598 | 280,6 |
| Black | 987.605 | 491 | 215,5 |
| Black Bag | 549.919 | 506 | 220,6 |
| Pool | 361.166 | 200 | 87,9 |
| Select | 1.713.733 | 340 | 170,6 |
| UberX | 2.233.720 | 377 | 277,9 |

Dados filtrados considerando apenas o produto UberX, uma grande amplitude nas médias e os desvios-padrão entre distritos podem ser vistos na Figura 3. É razoável considerar que um distrito com uma frota maior e mais acesso apresenta menor variação nos tempos de espera. Poderíamos dizer que os distritos mais acessíveis apresentam desvios-padrão mais baixos (Wang & Mu, 2018). Por esse motivo, dois modelos MLR foram criados. No primeiro, o tempo médio de espera é a variável dependente, e, no segundo, o desvio-padrão é a variável dependente.

Regressão linear múltipla (MLR)

Com base no princípio da estatística inferencial, é possível fazer afirmações sobre as características de uma população a partir de uma amostra dela. Análise de regressão é o termo que descreve uma família de métodos que permite explorar e inferir a relação entre duas ou mais variáveis (Francisco, 2010; Hair, 2006).

Figura 3. Comparação entre a média e o desvio-padrão dos tempos de espera estimados pelo UberX



Para a construção dos modelos MLR, foram utilizados apenas os dados do UberX agregados por distritos. A variável dependente para o modelo 1 foi o tempo médio de espera, e a variável dependente para o modelo 2 foi desvio-padrão. As variáveis independentes utilizadas foram:

- área (km²)
- população
- densidade populacional
- rendimento per capita – censo demográfico (em reais)
- empregos (comércio, serviços, indústria da transformação, construção civil)
- empregadores (comércio, serviços, indústria da transformação, construção civil)
- proporção de residentes não brancos (negros, pardos e indígenas)
- tempo de viagem
- taxa de motorização de automóveis domésticos e motocicletas
- número de paradas de ônibus
- comprimento das linhas de ônibus
- quantidade de linhas de ônibus

- número de estações de metrô

O software R e suas extensões “stats” e “car” foram utilizados para o cálculo dos modelos regressivos.

Modelo espacial autorregressivo (SAR)

Francisco (2010) sugere que, antes de criar um modelo SAR, seja verificada a autocorrelação espacial da variável dependente. A literatura utiliza uma medida estabelecida por Moran. O índice de Moran é um indicador da correlação entre o valor da variável observada em uma unidade espacial de análise e os valores dessa variável na região da unidade (seus vizinhos).

Após a verificação da autocorrelação geográfica dos tempos de espera pelo Uber por meio do I de Moran, foram selecionadas as variáveis altamente significativas do modelo MLR, e o SAR foi calculado. Francisco (2010) define SAR como um modelo de regressão capaz de incorporar a matriz dos vizinhos espaciais (ou proximidade espacial) como parte das variáveis explicativas.

O software GeoDa versão 1.12 foi usado para construir o SAR.

RESULTADOS

A Tabela 2 resume os resultados dos modelos MLR tendo como variável dependente a média dos tempos de espera pelo UberX e o desvio-padrão dos tempos de espera. É possível notar que o modelo de tempo médio possui maior poder de explicação, pois tem um R^2 de 0,893 contra um R^2 de 0,717 do modelo do desvio-padrão.

Tabela 2. Resultado das regressões de desvio médio e padrão para o UberX

| Variável dependente | | Média do UberX | | | Desvio-Padrão do UberX | | |
|---------------------|---|----------------|-------------|------------|------------------------|-------------|------------|
| | | Coefficiente | Erro-padrão | valor de p | Coefficiente | Erro-padrão | valor de p |
| (Intercepto) | | 943,00 | 750,40 | 0,213 | 39,35 | 515,30 | 0,939 |
| QTLINBUS2018 | Quantidade de linhas de ônibus | 0,13 | 0,19 | 0,514 | 0,12 | 0,13 | 0,353 |
| KMLINBUS2018 | Comprimento das linhas de ônibus (km) | -0,12 | 0,07 | 0,080 | -0,09 | 0,05 | 0,057 |
| QTPONTBUS2018 | Número de paradas de ônibus | -0,12 | 0,13 | 0,361 | 0,01 | 0,09 | 0,902 |
| QTESTMETRO2018 | Quantidade de estações de metrô | 0,91 | 5,22 | 0,862 | -1,14 | 3,58 | 0,751 |
| RENDP2010 | Rendimento <i>per capita</i> | -0,00 | 0,01 | 0,922 | 0,00 | 0,01 | 0,515 |
| ARE1 | Área (km ²) | 3,55 | 0,31 | 0,00 | 0,74 | 0,21 | 0,001 |
| POP2018 | População | 0,00 | 0,00 | 0,492 | -0,00 | 0,00 | 0,894 |
| DENPOP2018 | Densidade populacional | -0,00 | 0,00 | 0,007 | -0,01 | 0,00 | 0,000 |
| DOMP2018 | Número de domicílios particulares permanentes | -0,00 | 0,00 | 0,646 | 0,00 | 0,00 | 0,546 |
| ESTAB2016 | Empregadores | -0,00 | 0,01 | 0,970 | -0,00 | 0,00 | 0,791 |
| EMP2016 | Empregos | 0,00 | 0,00 | 0,384 | 0,00 | 0,00 | 0,547 |
| PNBRAN2010 | Proporção de não brancos | 300,20 | 77,69 | 0,000 | 202,20 | 53,35 | 0,000 |
| TEMP2010_5 MIN | Tempo de viagem – até 5 minutos | -587,50 | 1.166,00 | 0,616 | 111,00 | 800,50 | 0,890 |
| TEMP2010_30MIN | Tempo de viagem – de 6 a 30 minutos | -978,90 | 785,70 | 0,217 | 37,68 | 539,50 | 0,945 |
| TEMP2010_60MIN | Tempo de viagem – de 31 a 60 minutos | -278,40 | 831,70 | 0,739 | 180,80 | 571,10 | 0,752 |
| TEMP2010_120MIN | Tempo de viagem – de 61 a 120 minutos | -1,046, | 805,20 | 0,198 | -141,10 | 552,90 | 0,799 |
| TEMP2010_121MIN | Tempo de viagem – superior a 121 minutos | -1,443, | 992,80 | 0,150 | -215,90 | 681,80 | 0,752 |
| TEMP2010_oMIN | Viagem não realizada | -821,30 | 765,00 | 0,286 | 10,86 | 525,30 | 0,984 |
| PDOMC2010 | Carro – taxa de motorização doméstica | -139,10 | 367,70 | 0,706 | -164,10 | 252,50 | 0,518 |
| PDOMM2010 | Motocicleta – taxa de motorização doméstica | 37,73 | 82,42 | 0,648 | 25,17 | 56,60 | 0,658 |

Um modelo de MLR *stepwise* analisando a colinearidade das variáveis independentes do modelo final foi criado. Consideramos conveniente trabalhar com fatores de inflação de variância (FIVs) inferiores a 5, conforme recomendado por Batterham, Tolfrey e George (1997). Além disso, o modelo selecionado foi a média do tempo de espera, devido ao seu maior poder de explicação. A Tabela 3 mostra o resultado do modelo. Entretanto, a Tabela 4 mostra a análise de colinearidade das variáveis independentes selecionadas para esse modelo final. Podemos observar alta significância para o comprimento das linhas de ônibus, área do distrito, densidade populacional, empregos, porcentagem de não brancos e tempos de viagem superiores a 120 minutos, com um R-quadrado de 0,879. Esses elementos corroboram a visão positivista dos estudos de Lessa et al. (2019) e Páez et al. (2012).

Tabela 3. Resultados de regressão *stepwise* para tempos médios de espera pelo UberX

| Variável dependente | | Média do UberX | | |
|---------------------|--|----------------|-------------|------------|
| | | Coefficiente | Erro-padrão | valor de p |
| (Intercepto) | | 171,00 | 18,17 | 0,000 |
| KMLINBUS2018 | Comprimento das linhas de ônibus (km) | - 0,0786 | 0,02 | 0,002 |
| ARE1 | Área (km²) | 3,634 | 0,23 | 0,000 |
| DENPOP2018 | Densidade populacional | - 0,0035 | 0,0009 | 0,000 |
| EMP2016 | Empregos | 0,00021 | 0,0001 | 0,097 |
| PNBRAN2010 | Proporção de não brancos | 295,70 | 44,57 | 0,000 |
| TEMP2010_121MIN | Tempo de viagem – superior a 121 minutos | - 1,239,00 | 422,50 | 0,004 |

Tabela 4. Análise dos fatores de inflação de variação

| Fator de inflação de variância | | Média do UberX |
|--------------------------------|--|----------------|
| | | Coefficiente |
| KMLINBUS2018 | Comprimento das linhas de ônibus (Km) | 1,181254 |
| SÃO01 | Área (km²) | 2,139971 |
| DENPOP2018 | Densidade populacional | 1,317234 |
| EMP2016 | Empregos | 1,894758 |
| PNBRAN2010 | Proporção de não brancos | 2,462318 |
| TEMP2010_121MIN | Tempo de viagem – superior a 121 minutos | 2,615456 |

Figura 4. Tempo médio de espera pelo Uber X de acordo com o teste I de Moran

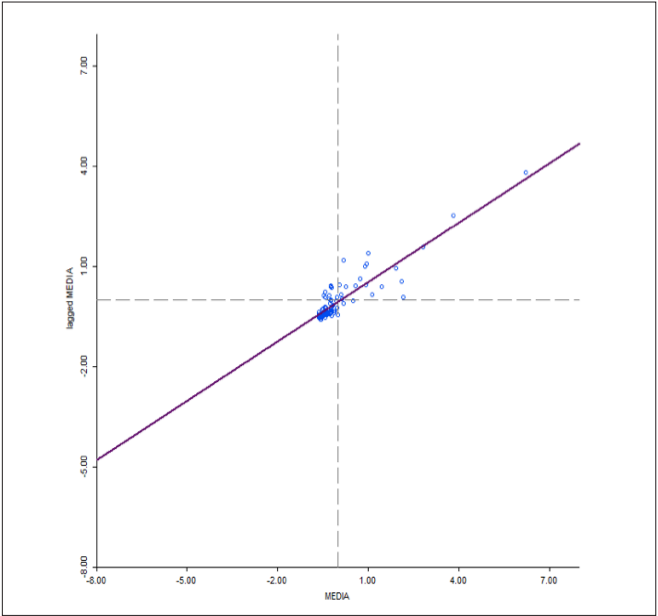
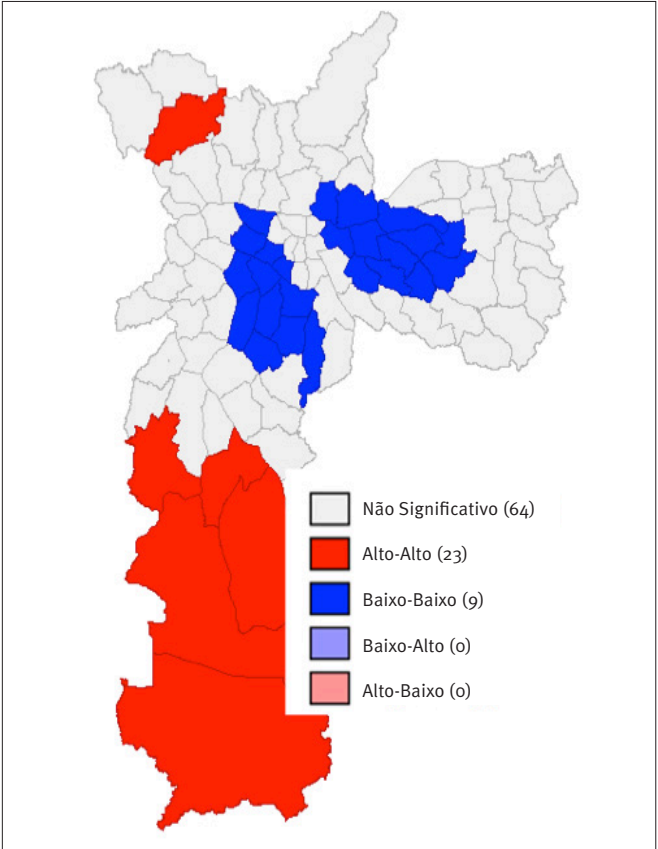


Figura 5. Tempo médio de espera pelo UberX por clusters de bairros



A Figura 4 mostra a dependência geográfica do tempo médio de espera pelo Uber X, com um I de Moran de 0,59, o que mostra a inclinação geográfica da variável dependente. Na Figura 5, observamos a distribuição da variável e sua dependência geográfica através do bairro, pois podemos notar áreas uniformes com menor tempo e baixa presença.

A partir das variáveis da Tabela 2, geramos um modelo de regressão espacial para o tempo médio de espera pelo UberX. Os resultados apresentados na Tabela 5 indicam alta significância para as variáveis comprimento das linhas de ônibus, área do distrito, densidade populacional e percentual de não brancos. Com a incorporação do fator geográfico no modelo, o R-quadrado salta para 0,89.

Tabela 5. Resultados da regressão de atraso espacial do UberX

| Variável dependente | | Média do UberX | | |
|---------------------|---------------------------------------|----------------|-------------|------------|
| | | Coefficiente | Erro-padrão | valor de p |
| W_MEDIA | | 0,406 | 0,092 | 0,000 |
| CONSTANTE | | 114,573 | 23,246 | 0,000 |
| KMLINBUS2018 | Comprimento das linhas de ônibus (km) | - 0,073 | 0,022 | 0,001 |
| ARE1 | Área (km ²) | 2,455 | 0,263 | 0,001 |
| DENPOP2018 | Densidade populacional | - 0,002 | 0,001 | 0,005 |
| PNBRAN2010 | Proporção de não brancos | 211,038 | 44,297 | 0,000 |

Observando os coeficientes do modelo SAR final, podemos dizer que o número de linhas de ônibus (-0,073) e a densidade populacional (-0,0023) influenciam negativamente o tempo de espera pelo Uber, assim como reportado por Wang e Mu (2018). A lógica é que a maior concentração de população e seu fluxo de transporte atrairiam os motoristas da Uber para essas regiões, impulsionando a oferta e proporcionando, assim, um menor tempo de espera para o serviço.

Em contrapartida, vale ressaltar que o coeficiente do percentual de não brancos (211,03) indica um aumento no tempo de espera quando a concentração de minorias é identificada na região. A possibilidade dessa relação – maiores tempos de viagem e espera *versus* distribuição de minorias – tem sido abordada de modo recorrente na literatura sobre acessibilidade (Flores & Rayle, 2017; Hughes & MacKenzie, 2016; Páez et al., 2012; Wang & Mu, 2018). Conforme sugerido por Wang e Mu (2018), as diferenças econômicas culturais

entre São Paulo e Atlanta podem ser o motivo da discrepância entre os resultados obtidos. Essa ocorrência nos motiva a continuar o debate sobre as idiossincrasias regionais de acessibilidade.

Ao analisar os resultados, não podemos confirmar H_1 , ou seja, não podemos afirmar que a estimativa do tempo de espera pelo serviço pode ser usada como proxy de acessibilidade. Isso se deve ao fato de que a relação com o tempo de espera pelo UberX não foi significativa para algumas variáveis.

Podemos confirmar H_2 para a distribuição da quantidade e número de linhas de ônibus, o que corrobora os resultados do estudo de Lessa et al. (2019), o qual encontrou que área do distrito, densidade populacional e o percentual de não brancos podem ser usados para medir acessibilidade. Nos estudos de Hughes e MacKenzie (2016) e Wang e Mu (2018), a relação entre o tempo de espera pelo Uber e distribuição de minorias não foi significativa.

DISCUSSÃO E CONCLUSÃO

O modelo final apresenta diferentes variáveis de estudos semelhantes. A taxa de minorias, por exemplo, apresenta uma elevada significância com um valor p de 0,00, em contraste com os casos de Seattle (Hughes & MacKenzie, 2016) e Atlanta (Wang & Mu, 2018). A relevância do contexto local é importante para explicar os achados deste estudo. Como os tempos de espera podem ter relação com a oferta e demanda consistentes de carros (Hall et al., 2019), uma explicação possível é que, como o padrão espacial socioeconômico (Torres et al., 2003) também acompanha as taxas de minorias e o status periférico, o serviço tende a ser menos acessível nessas regiões em que sua demanda poderia ser menor. No caso de São Paulo, a correlação entre taxas de minorias e renda nos encoraja a pelo menos pensar nessa possibilidade, já que a acessibilidade do Uber pode ser relativamente menor do que nos contextos de Seattle ou Atlanta. Uma futura análise usando dados das estimativas de custo da Uber pode ajudar a explorar esse padrão.

Por outro lado, algumas variáveis como densidade populacional (Hughes & MacKenzie, 2016; Lessa et al., 2019; Wang & Mu, 2018) e densidade de estradas (Wang & Mu, 2018), encontram alguma ressonância na discussão. Mesmo estando longe de ser suficiente para explicar a acessibilidade e estando cientes da natureza local do problema (Schwanen, 2017), podemos sugerir-las como variáveis que podem ajudar a descrever variáveis locais. Essa possibilidade existe porque a relação delas com a acessibilidade parece ser consistente.

A noção de que é necessário promover medidas de acessibilidade intuitivas e altamente replicáveis não está bem desenvolvida entre pesquisadores (Páez et al., 2012). Da mesma forma, a construção de uma medida de acessibilidade de transporte a partir de ferramentas de *Big Data*, como a API da Uber, pode contribuir para a divulgação e compreensão desse indicador para o público em geral na medida em que a teoria se aproxima de um serviço de consumo de massa. A expansão da mesma análise a outros prestadores de serviços e uma identificação mais precisa dessa participação de mercado no campo dos transportes em São Paulo também podem ajudar a identificar possíveis vieses no uso da ferramenta Uber.

A gestão pública pode fazer uso de ferramentas similares para desenvolver uma “cesta” de indicadores para diferentes modos de transporte e medições das suas inter-relações. Essas medidas podem ser utilizadas para regular melhor as atividades existentes e para adaptar a prestação de serviços públicos de mobilidade, com uma integração mais íntima entre os modos de transporte público e privado. Manter uma medida em tempo real do efeito da regulação pública em aspectos do sistema de transporte possibilitaria um melhor ajuste entre a intervenção e o interesse dos usuários (Jin et al., 2018), uma vez que indicadores bem-equilibrados podem refletir o comportamento e as necessidades dos usuários (Lessa et al., 2019; Páez et al., 2012).

No entanto, algumas limitações do estudo merecem ser destacadas, pois fornecem indicações para futuras pesquisas. É importante sublinhar a importância de uma análise mais abrangente. Este estudo limitou-se à frota da Uber. Em cidades como São Paulo, onde há pelo menos mais um ator com uma frota significativa, seria interessante replicar a metodologia com outras empresas que prestam serviços similares. Em segundo lugar, a replicação do estudo em outras cidades brasileiras parece necessária, pois os fatores econômicos e culturais que afetam as disparidades podem ser analisados em contextos mais semelhantes aos de Atlanta e Seattle, por exemplo. Em terceiro lugar, é necessário aprofundar a compreensão da ferramenta de estimativa de tempo da Uber para identificar possíveis vieses que poderiam ter suprimido alguma relação. Compará-la com outros serviços pode ser útil nesse sentido. Uma melhor compreensão desses dados pode representar um grande avanço em termos de ferramentas que possibilitem aos municípios atenderem melhor às necessidades de transporte de sua população. Além disso, vale a pena revisitar este estudo após a publicação do Censo do IBGE de 2020, pois muitas das variáveis utilizadas no estudo são projeções feitas a partir do Censo de 2010.

REFERÊNCIAS

- Aranha, V. (2005). *Mobilidade pendular na metrópole paulista. São Paulo em Perspectiva*, 19(4), 96-109. doi:10.1590/S0102-88392005000400006
- Batterham, A. M., Tolfrey, K., George, K. P. (1997). Nevill's explanation of Kleiber's 0.75 mass exponent: An artifact of collinearity problems in least squares models? *Journal of Applied Physiology*, 82(2), 693-697. doi:10.1152/jappl.1997.82.2.693
- Cohen, P., Metcalfe, R., Angrist, J., Chen, K., Doyle, J., Farber, H., Hahn, R. (2016). Using big data to estimate consumer surplus: The case of Uber. *National Bureau of Economic Research* (Working paper nº 22627). Retrieved from <https://www.nber.org/papers/w22627> doi: 10.3386/w22627
- Flores, O., & Rayle, L. (2017). How cities use regulation for innovation: The case of Uber, Lyft and Sidecar in San Francisco. *Transportation Research Procedia*, 25, 3756-3768. doi:10.1016/j.trpro.2017.05.232
- Francisco, E. de R. (2010). *Indicadores de renda baseados em consumo de energia elétrica: abordagens domiciliar e regional na perspectiva da estatística espacial* (Doctoral Thesis, Fundação Getúlio Vargas). Retrieved from <https://bibliotecadigital.fgv.br/dspace/handle/10438/8158>
- Hair, J., Babin, B., Money, A., & Samouel, P. (2005). *Fundamentos de métodos de pesquisa em administração*. Porto Alegre, RS: Bookman Companhia Ed.
- Hall, J. V, Horton, J. J., & Knoepfle, D. T. (2019). *Pricing efficiently in designed markets: The Case of Ride-Sharing*. 1-76. Retrieved from http://john-joseph-horton.com/papers/uber_price.pdf
- Hall, J. V, & Krueger, A. B. (2016). An analysis of the labor market for Uber's driver-partners in the United States. In *National Bureau of Economic Research* (Working paper nº 22843). Retrieved from <https://www.nber.org/papers/w22843>. doi:10.3386/w22843.
- Hansen, W. G. (1959). How accessibility shapes land use. *Journal of the American Institute of planners*, 25(2), 73-76. doi:10.1080/01944365908978307
- Hughes, R., & MacKenzie, D. (2016). Transportation network company wait times in Greater Seattle, and relationship to socioeconomic indicators. *Journal of Transport Geography*, 56, 36-44. doi:10.1016/j.jtrangeo.2016.08.014
- Instituto Brasileiro de Geografia e Estatística. (2010). Censo demográfico. *Notas Metodológicas-Microdados da Amostra. Rio de Janeiro*. Retrieved from ftp://ftp.ibge.gov.br/Censos/Censo_Demografico_2010/Resultados_Gerais_da_Amostra/Microdados/Documentacao.zip
- Jin, S. T., Kong, H., Wu, R., & Sui, D. Z. (2018). Ridesourcing, the sharing economy, and the future of cities. *Cities*, 76(January), 96-104. doi:10.1016/j.cities.2018.01.012
- Kim, B. G., Trimi, S., & Chung, J. (2014). Big-Data applications in the government sector. *Communications of the ACM*, 57(3), 78-85. doi:10.1145/2500873
- Krajewicz, D., Erdmann, J., Behrisch, M., & Bieker, L. (2012). Recent development and applications of SUMO-Simulation of Urban MObility. *International Journal On Advances in Systems and Measurements*, 5(3-4), 128-138.

- Kwan, M. (1998). Space-time and integral measures of individual accessibility: A comparative analysis using a point-based framework. *Geographical Analysis*, 30(3), 191-216. doi:10.1111/j.1538-4632.1998.tb00396.x
- Kwan, M. P. (2016). Algorithmic geographies: Big data, algorithmic uncertainty, and the production of geographic knowledge. *Annals of the American Association of Geographers*, 106(2), 274-282. doi:10.1080/00045608.2015.1117937
- Lessa, D. A., Lobo, C., & Cardoso, L. (2019). Accessibility and urban mobility by bus in Belo Horizonte / Minas Gerais – Brazil. *Journal of Transport Geography*, 77(May 2018), 1-10. doi:10.1016/j.jtrangeo.2019.04.004
- Letouzé, E., & Jütting, J. (2015). Official statistics, Big Data and human development: towards a new conceptual and operational approach. *Data Pop Alliance White Paper Series*. Retrieved from https://paris21.org/sites/default/files/WPS_OfficialStatistics_June2015.pdf
- Litman, T. (2003). Measuring transportation: Traffic mobility and accessibility. *Victoria Transport Policy Institute. ITE Journal*, 73(10), 28-32.
- McAfee, A., & Brynjolfsson, E. (2012). Big Data: The Management Revolution. *Harvard Business Review*. Retrieved from <https://hbr.org/>
- Metrô. (2008). Pesquisa Origem-Destino 2007. *Secretaria de Transportes Metropolitanos*. Retrieved from http://www.metro.sp.gov.br/pesquisa-od/arquivos/OD_2007_Sumario_de_Dados.pdf
- Ministério do Trabalho (2016). Relação Anual de Informações Sociais - RAIS ano-base 2016. *Portaria n. 1464 de 30 de Dez. 2016. Aprova Instruções Para a Declaração Da Relação Anual de Informações Sociais - RAIS Ano-Base 2016*. Retrieved from <http://www.rais.gov.br/sitio/sobre.jsf>
- Morandi, E., Ribeiro, R., Hernandez, E., Camara, B., Spinola, L., & Francisco, E. D. R. (2016). *Análise Geoespacial da Relação entre Transporte Público sobre Trilhos, Renda e Tempo Médio de Deslocamento*. Paper presented on XL EnANPAD, Salvador, BA. Retrieved from http://www.anpad.org.br/~anpad/eventos.php?cod_evento=1&cod_edicao_subsecao=1302&cod_evento_edicao=83&cod_edicao_trabalho=21023
- Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., & Mascolo, C. (2012). A Tale of many cities: Universal patterns in human urban mobility. *PloS One*, 7(9). doi:10.1371/annotation/ca85bf7a-7922-47d5-8bfb-bcdf25af8c72
- Páez, A., Scott, D. M., & Morency, C. (2012). Measuring accessibility: Positive and normative implementations of various accessibility indicators. *Journal of Transport Geography*, 25, 141-153. <https://doi.org/10.1016/j.jtrangeo.2012.03.016>
- Pääkkönen, P., & Pakkala, D. (2015). Reference architecture and classification of technologies, products and services for big data systems. *Big data research*, 2(4), 166-186. doi:10.1016/j.bdr.2015.01.001
- Preston, V., & McLafferty, S. (1999). Spatial mismatch research in the 1990s: Progress and potential. *Papers in Regional Science*, 78(4), 387-402. doi:10.1111/j.1435-5597.1999.tb00752.x
- Schwanen, T. (2016). Geographies of transport I: Reinventing a field? *Progress in Human Geography*, 40(1), 126-137. doi:10.1177/0309132514565725
- Schwanen, T. (2017). Geographies of transport II: Reconciling the general and the particular. *Progress in Human Geography*, 41(3), 355-364. doi:10.1177/0309132516628259
- Stelder, D. (2016). Regional accessibility trends in Europe: Road Infrastructure, 1957-2012. *Regional Studies*, 50(6), 983-995. doi:10.1080/00343404.2014.952721
- Torres, H. da G., Marques, E., Ferreira, M. P., & Bitar, S. (2003). Pobreza e espaço: Padrões de segregação em São Paulo. *Estudos Avançados*, 17(47), 97-128. doi:10.1590/S0103-40142003000100006
- Tribby, C. P., & Zandbergen, P. A. (2012). High-resolution spatio-temporal modeling of public transit accessibility. *Applied Geography Journal*, 34, 345-355. doi:10.1016/j.apgeog.2011.12.008
- Wang, M., & Mu, L. (2018). Spatial disparities of Uber accessibility: An exploratory analysis in Atlanta, USA. *Computers, Environment and Urban Systems*, 67, 169-175. doi:10.1016/j.compenvurbsys.2017.09.003
- Weibull, J. W. (1980). On the numerical measurement of accessibility. *Environment and Planning A: Economy and Space*, 12(1), 53-67. doi:10.1068/a120053
- Zhou, X., Wang, M., & Li, D. (2017). From stay to play—A travel planning tool based on crowdsourcing user-generated contents. *Applied Geography*, 78, 1-11. doi:10.1016/j.apgeog.2016.10.002