

Big Data e Urban Analytics à brasileira: questões inerentes a um país profundamente desigual

Flávia da Fonseca Feitosa 

Universidade Federal do ABC, Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas, Programa de Pós-Graduação em Planejamento e Gestão do Território, São Bernardo do Campo, SP, Brasil. Email: flavia.feitosa@ufabc.edu.br

<https://doi.org/10.47235/rmu.v8i1.141>

A crescente presença das Tecnologias da Informação e Comunicação (TICs) no cotidiano dos indivíduos vem provocando transformações profundas na sociedade, permeadas por novas formas de interação baseadas na intensa produção e uso de informações. Numa sociedade na qual a informação é matéria-prima essencial (Castells, 1999), o termo *big data*, ainda que associado a discursos frequentemente vagos e fragmentados, assumiu papel de destaque no âmbito empresarial, governamental e acadêmico. *Big data* remete, por si, à questão dos grandes volumes de dados estruturados e não estruturados constantemente produzidos na sociedade da informação.

Evidentemente, o potencial do *big data* só é atingido quando associado a processos que convertam grandes volumes de dados variados e velozes em informações relevantes, capazes de subsidiar tomadas de decisão. Tais processos demandam avanços não apenas no gerenciamento dos dados, mas também nas técnicas de análise. Consulta realizada na plataforma *Google Trends* revelou que o interesse mundial pelo termo *big data*, que vinha em ascendência desde o início da década de 2010, atingiu certa estabilidade a partir de 2017 (Figura 1). Concomitantemente, termos como *data science* e *data analytics* permanecem em

trajetória ascendente, tendo o termo *data science*, inclusive, despertado mais interesse do que *big data* desde 2018.

Estaria o entusiasmo com *data science* e *data analytics* associado a novas roupagens para técnicas já estabelecidas em análise de dados? Em parte. Cabe, entretanto, reconhecer que os desafios inerentes à análise dos dados produzidos nos dias atuais, cada vez mais volumosos e complexos, vêm atribuindo a esses termos um significado fortemente associado à interdisciplinaridade, que combina campos como os da matemática, da estatística e da ciência da computação.

No campo específico do planejamento e gestão urbana, os termos *smart cities* e *urban analytics* ganharam destaque, embora o primeiro venha despertando ceticismo no meio acadêmico da área de planejamento. Söderström et al. (2014) destacam como o conceito de *smart cities* foi mobilizado em narrativas delineadas por grandes corporações de tecnologia da informação (TI). Para os autores, tais narrativas concebem a cidade como um sistema de sistemas e reciclam um discurso utópico voltado a apontar patologias urbanas bem como sua “cura”, que, por sua vez, envolve altos investimentos em TI e fortalece a posição dessas corporações no mercado.

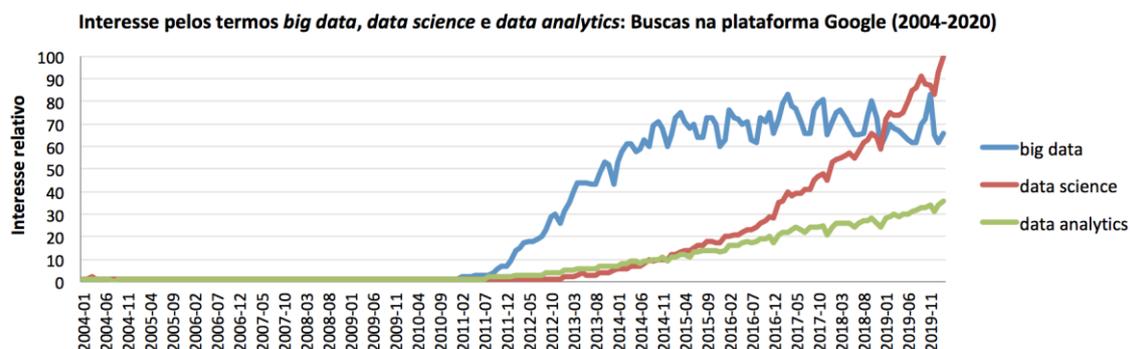


Figura 1. Interesse pelos termos *big data*, *data science* e *data analytics* segundo consultas realizadas na plataforma *Google* no período de 2004 a 2020 (fonte: *Google Trends*, 02/03/2020).

Em todo caso, independente das críticas relacionadas ao caráter comercial, as iniciativas *smart* permanecem, na prática, mais associadas ao gerenciamento de serviços urbanos e a demandas de curto prazo.

A *Urban analytics*, por sua vez, vem despontando como um campo emergente no ambiente acadêmico, fortemente associado a pesquisas que promovam olhares inovadores sobre fenômenos e dinâmicas urbanas. O entusiasmo reflete-se no surgimento recente de inúmeros programas de pós-graduação em *urban analytics* no Reino Unido (*University of Glasgow, University College London, University of Warwick, Manchester Metropolitan University*), Estados Unidos (*New York University, Northeastern University*) e Hong Kong (*University of Hong Kong*).

Goodchild, em citação incluída nas páginas iniciais do livro *Urban Analytics* de Singleton e colegas, relaciona o campo a uma nova modalidade de pesquisa urbana, que “explora os novos vastos recursos de dados que estão se tornando disponíveis, nas mídias sociais, *crowdsourcing* e redes de sensores; e utiliza o poder sem precedentes da atual tecnologia computacional” (Singleton et al., 2018, sem página, tradução nossa). Batty (2019), entretanto, flexibiliza a relação entre *urban analytics* e *big data*, afirmando haver espaço para buscar novos sentidos e formas de explorar dados que não são volumosos e nem gerados em tempo real (*small data*). O autor flexibiliza também a própria noção de *big data*, ao sugerir que dados tradicionais, coletados por meio de questionários, podem ser tratados como *big data* sob certas condições, como, por exemplo, em forma desagregada ou combinados com outros dados (Batty, 2016; 2019).

Para Batty (2019), a questão-chave a ser tratada pelo campo de *urban analytics*, a qual representa o grande desafio da área, consiste em avançar para além da análise de dados em si, rumo à construção de novas teorias urbanas – ou nova ciência da cidade – e estratégias para aprimorar o planejamento das cidades. Esse processo deverá, sim, incluir os novos dados, bem como os modelos e métodos pesquisados sob a rubrica da *urban analytics*, mas sempre servindo de base para nos auxiliar a entender, explicar e tratar grandes questões contemporâneas (Batty, 2019).

É relevante que essa reflexão seja realizada também em relação ao contexto brasileiro. A chamada “revolução dos dados” é perceptível em nossas cidades. Mas como ela vem ocorrendo? Qual sua abrangência? De que forma vem avançando rumo a uma melhor compreensão e planejamento das nossas áreas urbanas? Como pode contribuir para entender e tratar de questões cruciais de nossa realidade, dentre as quais se destacam as acentuadas desigualdades em suas múltiplas dimensões?

Big data e Urban Analytics nas cidades brasileiras: novas formas de pensar o urbano? Qual urbano?

O Brasil também tem participado com interesse do debate em torno das transformações promovidas pelas TICs, bem como as possibilidades do *big data* e termos relacionados. Na cidade do Rio de Janeiro, por exemplo, o ideal comercial de *smartificação* de cidades alcançou um de seus maiores feitos com a construção do Centro de Operações Rio (COR). O COR representa um exemplo de solução *smart* na qual informações em tempo real, provenientes de câmeras e sensores sobre as condições de vários serviços e infraestruturas urbanas, são utilizadas para o monitoramento e otimização do funcionamento da cidade, bem como emissão de alertas e deflagração de protocolos de emergência em situações críticas. As ações de monitoramento e gestão de serviços são complementadas por iniciativas voltadas à comunicação com o público em geral, que vão desde uma sala reservada à imprensa até aplicativos com informações sobre previsão do tempo, transporte e trânsito.

Enaltecimentos e críticas envolvendo as mais distintas perspectivas têm sido realizados em relação à experiência. Cabe, entretanto, salientar uma limitação relevante relacionada ao tema aqui discutido: a excessiva ênfase ao monitoramento, planejamento e ações de curto prazo.

Trata-se de uma observação que não se restringe à realidade do COR e do Rio de Janeiro, tendo sido pertinente a inúmeros outros contextos. Para Batty (2019), o avanço do *big data* tem, de fato, promovido análises de curto prazo sobre como as cidades funcionam e podem ser gerenciadas, em detrimento de um planejamento estratégico de longo prazo. O autor afirma, entretanto,

que esses dados podem se tornar fontes de informação para qualquer horizonte de tempo, desde que considerados períodos muito mais longos de tempo.

No caso específico do COR, a predominância de uma visão de curto prazo reflete-se também nas restrições existentes para armazenamento dos dados, realizado por tempo limitado devido ao grande volume de informações gerado. A incapacidade de armazenar e sistematizar os dados produzidos por longos períodos de tempo restringe estudos futuros voltados à análise de longas séries temporais e desperdiça o real potencial do *big data* e *urban analytics* para um avanço efetivo na compreensão das dinâmicas de nossas cidades e para o planejamento de médio e longo prazo.

Complementarmente, devemos ressaltar a relevância de consolidação de uma política efetiva de abertura dos dados para fins de inovação, seja no âmbito acadêmico, governamental ou da iniciativa privada. Tal política envolve estabelecer uma participação cidadã não apenas na produção e compartilhamento de dados, mas também em sua apropriação e utilização de forma ativa. Algumas iniciativas nesse sentido foram realizadas, incluindo *hackatons*, *living labs* e, mais recentemente, o lançamento do Projeto labGov.RIO, que conta com o apoio do COR, e prevê a construção de um centro de inovação voltado para empreendedores, pesquisadores, desenvolvedores e investidores na área de tecnologia e inovação.

Cabe verificar, entretanto, como e se essas iniciativas vêm atuando positivamente na descentralização do “pensar a cidade” para fins de planejamento. É fundamental que se construam oportunidades de potencializar o retorno dos investimentos públicos realizados na construção e manutenção desse aparato tecnológico na forma de novos conhecimentos sobre as dinâmicas e padrões da cidade bem como efetivas avaliações das ações públicas e seus reflexos sobre a cidade. Assim, será possível subsidiar a elaboração de planos e políticas urbanas bem informadas e com horizontes para além do curto prazo.

Inúmeras questões intersetoriais poderiam ser avaliadas a partir do acesso a séries temporais dos dados produzidos pelo COR. Partindo de uma perspectiva voltada à segurança pública, por exemplo, seria

possível avaliar questões como: considerando distintas escalas temporais (dia/noite, meses e anos), como o perfil das ocorrências policiais vem se modificando territorialmente ao longo do tempo? De que forma vem ocorrendo o atendimento por parte do poder público? Como essas dinâmicas relacionam-se com a evolução de indicadores socioeconômicos, com as transformações na ocupação do território (perfil sociodemográfico, uso do solo etc.) e com o histórico das políticas de uso e ocupação do solo, mobilidade, educação, saúde e segurança? Para quais cenários e desafios devemos nos preparar e planejar?

De maneira geral, a ampla disponibilização dos *big data* - mais especificamente, a conversão de *big data* em *open data* - representa uma questão-chave e urgente a ser tratada. Embora, legalmente, os dados governamentais sejam públicos, sabe-se que, na prática, há muito a se avançar na efetivação de políticas de disponibilização.

Cabe ressaltar, ainda, que parcela significativa dos grandes dados gerados na atualidade não estão sob domínio de instituições públicas, mas sim privadas. O aumento da relevância dos *big data* tendem a aumentar o protagonismo da iniciativa privada na produção de dados, o que nos leva a maiores dificuldades de abertura dos dados e incertezas sobre as finalidades dos usos dos mesmos. A abertura dos *big data* representa um passo fundamental para compartilhar o poder atribuído aos detentores de dados e informações, o que, por conseguinte, também implica na democratização e potencialização de seu uso.

Outro problema recorrente, também relacionado à disponibilização dos *big data*, diz respeito à fragmentação e descontinuidade de sua cobertura, frequentemente restrita a limites político-administrativos. Assim como a infraestrutura do COR, digna de países desenvolvidos, restringe-se ao Município do Rio de Janeiro, muitos dos novos dados disponibilizados não permitem uma análise do espaço regional ou mesmo do espaço intraurbano (Villaça, 1998), marcado pelos deslocamentos cotidianos de pessoas, seja como força de trabalho, seja como consumidores.

Na Região Metropolitana de São Paulo, por exemplo, os avanços na disponibilização de dados para estudos de mobilidade e

acessibilidade concentram-se principalmente no Município de São Paulo. Dados de transporte público no formato *General Transit Feed Specification* (GTFS), por exemplo, estão disponíveis apenas para a capital e algumas linhas metropolitanas de ônibus, mas não para as linhas locais dos demais municípios que compõem a região metropolitana. Esse fato induz a uma concentração de estudos voltados para apenas uma porção da metrópole - a mais privilegiada - e inviabiliza análises que explorem o potencial dos novos dados para avançar no entendimento e planejamento da mobilidade na metrópole como um todo.

Se tais desigualdades de informações são observadas no interior da maior e mais rica região metropolitana do país, o que dizer sobre as cidades mais afastadas dos grandes centros? Essa questão é particularmente relevante em um Brasil de tantos “Brasis”, muitos deles pouco conhecidos, e que já tradicionalmente desenha suas políticas a partir da perspectiva dos grandes centros.

O entusiasmo diante dos novos dados tende a motivar pesquisas restritas ao tipo de dado disponível e às regiões onde estão disponíveis, que não necessariamente dialogam com uma necessidade ou problema estabelecido *a priori*. Tal debate nos leva a refletir sobre limitações práticas que enfrentamos na utilização do *big data* e *urban analytics* para a compreensão da natureza do urbano no nosso país. Afinal, *urban analytics* para qual urbano?

A visão de urbano, assim como de sociedade, implícita no *big data* é necessariamente restrita. Conforme ressalta Shearmur (2015), ao lidarmos com *big data* não estamos tratando da sociedade, mas de usuários e mercados - os usuários de *twitter* ou *facebook*, usuários de transporte coletivo, motoristas conectados a aplicativos de navegação por GPS, clientes de concessionárias de telecomunicações, energia, água e esgoto etc. São dados tendenciosos por natureza, que não dizem respeito às pessoas que estão alheias aos mercados e atividades específicas que estão sendo rastreadas.

É certo que também os países desenvolvidos são afetados pelas questões aqui levantadas: amostragem enviesada, com visão restrita a usuários e mercados; acesso ainda limitado aos dados, frequentemente sob domínio de

empresas privadas; bem como abrangência geográfica fragmentada e comumente restrita a limites político administrativos. Entretanto, diante das profundas desigualdades de um país como o Brasil, em suas distintas dimensões e escalas, esse debate merece protagonismo.

Estaríamos contribuindo para acentuar as disparidades entre aqueles espaços que Milton Santos retratou como luminosos e opacos (Santos, 1996)? Estaríamos voltando nosso olhar a determinados mercados que excluem boa parte da sociedade? Estaríamos contribuindo para tornar invisíveis aqueles que mais precisam de visibilidade? Estaríamos tentados a voltar nossas atenções a problemas que os novos dados e técnicas analíticas a eles associados permitem explorar, em detrimento de questões mais cruciais para o bem estar da sociedade?

O papel dos dados tradicionais na era do *big data*

Em 2015, o Conselho da Agenda Global do Fórum Econômico Mundial sobre o Futuro do *Software* e da Sociedade divulgou uma pesquisa realizada com cerca de 800 líderes setoriais a respeito de 21 pontos de inflexão que deverão ser proporcionados por mudanças tecnológicas até 2025. Dentre esses pontos de inflexão, chama atenção o que trata do “primeiro governo a substituir seu censo por fontes de *big data*”. A pesquisa revela que 83% dos entrevistados acreditavam que isso ocorreria antes de 2025 (World Economic Forum, 2015).

Esse resultado ilustra como a discussão sobre o papel dos levantamentos tradicionais na era do *big data* frequentemente tem assumido contornos que induzem à diminuição de sua importância ou mesmo sua extinção.

No decorrer de 2019, o Brasil assistiu a um intenso debate sobre os cortes orçamentários do censo demográfico do país. De maneira geral, o Governo Federal insistiu em discursos voltados à desvalorização dos levantamentos tradicionais e desqualificação das instituições que os produzem. No caso específico do censo, as justificativas dos defensores da redução do orçamento e do questionário a ser aplicado variavam. Enquanto a frase “*quem pergunta demais descobre o que não quer*”, proferida pelo Ministro da Economia, escancarava o descomprometimento com a realidade

brasileira e com a elaboração de políticas públicas bem informadas, justificativas aparentemente mais técnicas focavam no *big data* e registros municipais como formas alternativas de coleta de informações.

Como resultado, foram suprimidas do censo demográfico 2020 questões relacionadas ao custo da moradia, rendimento, emigração, bem como acesso a serviços e posse de bens. A supressão das questões foi efetuada às vésperas da realização do levantamento, sem qualquer debate com a comunidade interessada ou mesmo justificativa para a seleção dos temas sobre os quais passaremos a ter menos informações.

Em relação aos argumentos baseados no uso de outras fontes de dados, também não houve qualquer esclarecimento sobre como as informações suprimidas no censo poderiam ser substituídas. O fato é que não temos alternativas para suprir, com qualidade, essa lacuna.

Por exemplo, no caso da variável “custo do aluguel”, utilizada nas estimativas do déficit habitacional, sabe-se que dados sobre preços dos imóveis são monitorados constantemente por empresas privadas e utilizados para estudos de interesse do mercado imobiliário. Esses dados, além de privados, são restritos ao mercado imobiliário formal de algumas regiões do país. Sua qualidade costuma ser muito superior a dos dados municipais que, além de também possuírem abrangência restrita, são frequentemente desatualizados e desconectados com a realidade dos municípios. Municípios estes que, em muitos casos, não possuem sequer informações confiáveis sobre a situação fundiária de boa parte de seu território.

Qual seria, portanto, a alternativa à questão suprimida no questionário do censo? Diante da inexistência de tal alternativa, resta aos municípios brasileiros ficar às cegas em relação a uma das mais expressivas dimensões do déficit habitacional, o custo excessivo da moradia, que atingia mais de 2,1 milhões de domicílios de baixa renda em 2010 (Fundação João Pinheiro, 2013).

Referências

Batty, M. (2016) Editorial: Big Data, Cities and Herodotus. *Built Environment*. 42(3),

As fontes de *big data* abrem possibilidades para explorarmos novos padrões e dinâmicas até então desconhecidos, mas ainda são dados de abrangência restrita e tendenciosos por natureza. É importante associá-los, portanto, a dados levantados por meio de rigoroso planejamento e desenho amostral, tais como o censo demográfico.

Além de ampla cobertura do território, o censo demográfico dispõe de um rico espectro de variáveis, estabelecido a partir de amplo debate com a comunidade de usuários, com claras definições de conceitos e categorias, e em sintonia com questões relevantes para o país.

O censo oferece, portanto, uma base confiável para calibração e exploração dos novos dados, bem como variáveis relevantes que potencializam a interpretação de resultados. Por outro lado, a análise de alguns fenômenos e dinâmicas tradicionalmente observados por meio dos dados censitários pode ser conduzida nos períodos intercensitários com auxílio de dados alternativos, que são constantemente produzidos. Por exemplo, análises sobre dinâmicas e tendências relacionadas ao poder de compra dos brasileiros podem combinar dados censitários, levantados a cada dez anos, com dados de consumo de energia elétrica, gerados constantemente.

É fundamental, portanto, um debate sobre o papel dos levantamentos tradicionais na era do *big data* e *urban analytics*. A exploração de estratégias de integração entre dados tradicionais e *big data* torna-se crucial para potencializar o uso de ambos em análises que contribuam para avançar na construção de novas teorias urbanas e aprimoramento de planos e políticas urbanas. Não se pode admitir que o entusiasmo com o *big data* subsidie discursos e ações que depreciem e degenerem a produção de dados sobre o país. Movimentos nessa direção atendem apenas a interesses de governos autoritários, avessos à transparência, para os quais dados sobre a realidade do país não passam de um inconveniente.

317-320. Disponível em: <https://doi.org/10.2148/benv.42.3.317>.

Batty, M. (2019) Urban Analytics Defined. *Environment and Planning B: Urban*

- Analytics and City Science*. 46(3), 403-405. Disponível em: <https://doi.org/10.1177/2399808319839494>.
- Castells, M. (1999) *A sociedade em rede*. São Paulo, Paz e Terra.
- Fundação João Pinheiro. Centro de Estatística e Informações (2013) *Déficit habitacional municipal no Brasil 2010*. Belo Horizonte, Fundação João Pinheiro. Disponível em: <http://www.bibliotecadigital.mg.gov.br/consulta/consultaDetalheDocumento.php?iCodDocumento=73954>. [Consultado em: 25 de junho de 2020].
- Santos, M. (1996) *A natureza do espaço: espaço e tempo: razão e emoção*. São Paulo, Hucitec.
- Shearmur, R. (2015) Dazzled by Data: Big Data, the Census and Urban Geography. *Urban Geography*. 36(7), 965-968.
- Disponível em: <https://doi.org/10.1080/02723638.2015.1050922>.
- Singleton, A., Spielman, S., Folch, D. (2018) *Urban Analytics*. London, Sage.
- Söderström, O., Paasche, T. e Klauser, F. (2014) Smart Cities as Corporate Storytelling. *City*. 18(3), 307-320. Disponível em: <https://doi.org/10.1080/13604813.2014.906716>.
- Villaça, F. (1998) *Espaço intra-urbano no Brasil*. São Paulo, Studio Nobel.
- World Economic Forum (2015). *Deep Shift - Technology Tipping Points and Societal Impact*. Survey Report. Disponível em: http://www3.weforum.org/docs/WEF_GAC15_Technological_Tipping_Points_report_2015.pdf. [Consultado em : 27 de fevereiro de 2020]