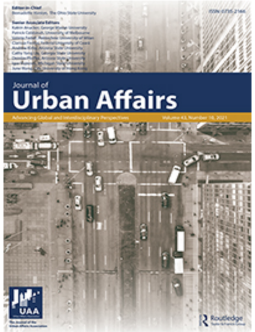

Journal of Urban Affairs

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ujua20

Reconciling big data and thick data to advance the new urban science and smart city governance

Andy Hong, Lucy Baker, Rafael Prieto Curiel, James Duminy, Bhawani Buswala, ChengHe Guan & Divya Ravindranath

To cite this article: Andy Hong, Lucy Baker, Rafael Prieto Curiel, James Duminy, Bhawani Buswala, ChengHe Guan & Divya Ravindranath (2022): Reconciling big data and thick data to advance the new urban science and smart city governance, Journal of Urban Affairs, DOI: 10.1080/07352166.2021.2021085

To link to this article: https://doi.org/10.1080/07352166.2021.2021085

JOURNAL OF URBAN AFFAIRS
https://doi.org/10.1080/07352166.2021.2021085

Reconciling big data and thick data to advance the new urban science and smart city governance

Andy Hong[a,b], Lucy Baker[b], Rafael Prieto Curiel[c], James Duminy[d,e], Bhawani Buswala[b], ChengHe Guan[f], and Divya Ravindranath[g]

[a]University of Utah; [b]University of Oxford; [c]University College London; [d]University of Bristol; [e]University of Cape Town; [f]New York University Shanghai; [g]Indian Institute for Human Settlements

ABSTRACT
Amid growing enthusiasm for a ”new urban science” and ”smart city” approaches to urban management, ”big data” is expected to create radical new opportunities for urban research and practice. Meanwhile, anthropologists, sociologists, and human geographers, among others, generate highly contextualized and nuanced data, sometimes referred to as ‘thick data,’ that can potentially complement, refine and calibrate big data analytics while generating new interpretations of the city through diverse forms of reasoning. While researchers in a range of fields have begun to consider such questions, scholars of urban affairs have not yet engaged in these discussions. The article explores how ethnographic research could be reconciled with big data-driven inquiry into urban phenomena. We orient our critical reflections around an illustrative example: road safety in Mexico City. We argue that big and thick data can be reconciled in and through three stages of the research process: research formulation, data collection and analysis, and research output and knowledge representation.

KEYWORDS
Big data; thick data; urban science; smart cities; road safety; ethnography

Introduction

In concert with the emergence of a “new urban science” (Acuto, 2018; Acuto et al., 2018; Batty, 2012; Karvonen et al., 2021; Ortman et al., 2020) and “smart” approaches to city management (Batty et al., 2012; Coletta et al., 2018; Dameri et al., 2019), there is growing enthusiasm for the availability of “big data” to create new opportunities for innovative and effective modes of urban analysis and governance. High-powered computational capability coupled with new analytic techniques employing artificial intelligence have created a tipping point in our capacity to deal with massive amounts of continuously generated data that are exhaustive and highly granular in both temporal and spatial scales (Kitchin, 2014a). The term *big data* refers to massive data sets that may be analyzed computationally to reveal associations, patterns, and trends, but that also entail varied and complex structures posing challenges of storage, analysis, and visualization (Sagiroglu & Sinanc, 2013). It can be characterized as data with three Vs: volume, variety, and velocity (Ahn et al., 2013).

These new forms of data analysis are increasingly applied to a broad range of research topics within the social sciences (Cinnamon et al., 2016; Rubrichi et al., 2018; Salah et al., 2018; Wilson et al., 2016; Zufiria et al., 2018). For urban researchers and professionals, the unprecedented availability of geotagged, individualized data collected through a multitude of devices distributed throughout city spaces can potentially provide new and integrated insights into the experience and evolution of cities through efficient real-time monitoring and algorithmic applications (Li et al., 2020; Rabari & Storper, 2015). That heightened interest is reflected in the creation of the journal *Urban Science* and the renaming of

CONTACT James Duminy james.duminy@bristol.ac.uk School of Geographical Sciences, University of Bristol, University Road, Bristol BS8 1SS, UK.

Environment and Planning B with the subtitle *Urban Analytics and City Science*. Moreover, big data technologies have underpinned new forms of urban planning and management practice represented by the rolling out of "smart cities" and "urban laboratories" (Hu & Zheng, 2021; Karvonen & Van Heur, 2014; Smith et al., 2019; Tricarico et al., 2020). These modes of city governance express a normative epistemology that defines the city as a space of experimentation; a place to be known "as a set of variables" and "a collection of parameters that can be tinkered with and controlled" (Barnett, 2020; Caprotti & Cowley, 2017).

Big data can potentially be the source of significant advances in our understanding and management of urban systems and dynamics. However, as we will show, such data may struggle to capture more nuanced and contextualized forms of knowledge about diverse urban places or the people that reside, move, and work therein. Big data are typically derived from technologies that automatically generate large data sets, collected by various sensors and mobile devices—practices that are seldom analyzed on established epistemological or theoretical grounds (Kitchin, 2014a). Thus, while offering unprecedented opportunities for prediction and projection, big data analytics may be less suitable for exploring or interpreting causal relations in the data in order to make inferences relevant to complex urban questions, including what effects a certain intervention might create, and the unintended effects that could result (Athey, 2017).

While a number of authors refer to human "smartness" to describe the capital, entrepreneurship, and innovation that engage citizens in participatory and collaborative processes of the city (Kitchin, 2014b; De Oliveira et al., 2015; Rabari & Storper, 2015), the strengths and advantages of 'thick' data have not been a focus for the ways in which the smart city is imagined or practiced. The potential of thick data, including the insights and interpretations emerging from ethnographic methods, lies in the capacity to provide more nuanced and contextualized information that would be particularly relevant for urban places and societies that are not adequately captured or examined by the practices typically associated with big data analytics (Bornakke & Due, 2018; Girardin, 2013; Wang, 2013).

As we will show, scholars from a range of disciplines and fields have argued for the reconciliation of big data analytics and ethnographic inquiry as a means to generate new kinds of knowledge processes and insights. However, to date urban scholars have been absent in these discussions, with several results. One is that connections between ethnographic methods and the use of digital technologies, including social media, remain undertheorized in urban research (Danley, 2021). A second is that calls for the combination of "cutting-edge" methodologies in urban research, such as ethnography and advanced spatial analysis, have largely remained unanswered (Reese, 2014). And third, the notion that urban scholars—with their extensive experience in conducting interdisciplinary research—might have something productive to say in broader methodological discussions surrounding big and thick data has gone unrecognized. Consequently, we seek to contribute to emerging interdisciplinary engagements surrounding big and thick data from a specifically urban perspective. Moreover, by identifying the parameters of the problem and exploring the scope for methodological and epistemic reconciliation, this discussion contributes to debates within the field of urban analytics emphasizing that the "new urban science" must necessarily comprise more than an exclusive focus on data analytics. This is imperative if those working within the field are to capture adequately the long-term temporalities of urban change, to decisively influence urban governance practice, and to address places and populations typically excluded from the purview of big data (Batty, 2019; Duminy & Parnell, 2020; Kang et al., 2019; Karvonen et al., 2021; Keith et al., 2020).

In this paper we argue in favor of a "hybrid" approach and explore initial ideas for how big data could potentially be combined with the contextual and descriptive depth of thick data to enrich our understanding of the complex nature of contemporary urban change. We have used an illustrative example (road traffic safety in Mexico City) to prompt review-based inquiry and discussion to illustrate how a hybridized approach would help improve decision-making by integrating big and thick data for the benefit of the public sector and civil society. Through our interdisciplinary approach, we focus on the specific question of how ethnographic methods and knowledge could potentially complement, refine, and extend the examination of urban phenomena driven by big data analytics.

The insights generated through this approach can assist actors involved in urban governance to leverage big data by integrating various disciplinary expertise, research objectives, analytical outcomes, and forms of reasoning made possible by methodologies and sensibilities related to both big and thick data.

Moreover, the paper provides an example of how interdisciplinary research teams, including both quantitative and qualitative urban experts, can be built as well as how diverse stakeholders can be engaged in collecting and analyzing data. It is, thus, our intention to push the emerging experimental and "smart" fields of urban governance to improve their accuracy in the analysis of big data to interpret and address complex urban problems while moving between more or less "generalizable" and "contextualized" forms of knowledge.

Results: Reviewing the emerging fields of big data and thick data

We began our assessment of this emerging field by reviewing existing research that combines big and thick data and their related methods. While a mixed-method approach could potentially break down disciplinary silos surrounding big and thick data methods, we found that mixed-method analyses were uncommon in studies employing big data. Most of the mixed-method approaches we observed utilize qualitative data collected through interviews, focus groups, and surveys. A limited number of studies, however, have specifically engaged with ethnographic methods as a complement to big data methods. For this reason, and because our research team comprises both big data scientists and researchers with ethnographic expertise, we focused on reviewing a number of carefully chosen studies that combine thick data arising from ethnographic methods with big data-driven computational analytics. We did this in order to establish current trends in the way that ethnographic data and methods are being employed in conjunction with big data and to thus identify new and improved ways of reconciling big and thick data specifically for the study, planning, and management of cities. In the following section, we report some of the key points and insights from our review of the literature.

Acknowledging the limitations of big data–driven urban analytics

To illustrate the respective dimensions, strengths, and limitations of big data and thick data, we provide a simplified comparison of these forms in Table 1. One of the key advantages of big data-driven research is the capacity to generate information at unprecedented scale, granulation and depth on urban phenomena ranging from workplace segregation (Zhou et al., 2021), real estate pricing (Cao et al., 2019; Tan & Guan, 2021), local social organizational change (Wang & Vermeulen, 2021), the dynamics of park usage (Guan et al., 2020, 2021; Li & Yang, 2021), and how transport networks affect inter-city interaction and economic integration (Gong et al., 2021). During the ongoing COVID-19 pandemic, for example, sources of big data and advanced analytical techniques have proven useful in examining public sentiments toward public health governance measures (Yao et al., 2021) and the impact of the urban built environment and infrastructure access on disease incidence and vulnerability (Kawlra & Sakamoto, 2021; Liu et al., 2021).

While such analyses are critical for producing insights into city dynamics at large scales, it is also possible that extremely large data sets can produce "too much information, with no easily specifiable social context" or "ground truth" (Blok et al., 2017). Consequently, collecting and analyzing big data sets alone may not be sufficient to give comprehensive answers to questions relating to complex human behaviors (Girardin, 2013; Kim et al., 2018). Such data may also be too abstract to give insights into the emergence and influence of place-based meanings and practices, unless it can be enlivened to narrate a story that "makes sense" (Dourish & Gómez Cruz, 2018). Thus, big data analyses have been critiqued for their lack of a contextual framing (Boyd & Crawford, 2012; Van Dijck, 2014). Furthermore, it cannot be assumed that big data, although extensive in reach, is necessarily representative of all population groups and territorial spaces (Boy et al., 2016) and in some instances, as with data extracted from social media, socioeconomic information is not available (Kim et al., 2018).

Table 1. Comparison of big data and thick data.

	Big data	Thick data	Hybrid approach
Definition	Massive data sets that can be analyzed computationally to reveal associations, patterns, and trends	Qualitative information collected and analyzed to develop a holistic view of a phenomenon or practice, including relevant context, how people construct meanings, and why they act in certain ways	An approach that seeks to reconcile big data analytics with ethnographic research practices and thick data
Types of data	Often quantitative and machine-readable	Qualitative and not easily machine-readable	Flexible data types according to a given situation
Data size and scale	Large size and spatial scale often requiring high throughput computing power	Small size and limited spatial scale, but often rich and in-depth	Multi-scalar and diverse
Data collection methods	Large-scale surveys, administrative data, digital archives, data streaming, web scraping, remote sensing, digital devices	Interviews, focus groups, video and audio recordings, field observations, narrative accounts, community/participant mapping, archival analysis, drawing, and photography	Big data-aided ethnography, highly granular spatiotemporal data from sensors and ethnographic fieldwork, calibrated machine learning analytics
Analytic methods	Mathematical models, agents, statistical methods, artificial intelligence, machine learning, computational modeling, etc.	Content analysis, interpretive thematic analysis, grounded theory, discourse analysis, narrative analysis, case study and comparative case analysis, etc.	Flexible methods based on available data
Representation of knowl-edge	Statistical relationships, hypothesis testing, and numerical or spatial visualizations	Written, visual, oral or performed narrative accounts	Diversified forms of representation
Output forms	Short-form and image intensive journal articles, maps, animations, videos, computational models, algorithms, equations	Longer-form journal articles and book chapters, monographs, performances, sketches, films, stories, photographic essays, community/stakeholder working groups	Diverse outputs integrating visualization and narration, e.g. story mapping
Strengths	Produces results that are empirically and statistically generalizable and comparable to a larger population	Produces results that uncover meaning, lived experience, and context, including those of marginalized populations, through targeted sampling	Produces actionable insights that are nuanced, contextualized, and statistically generalizable
Limitations	Lacks depth and key contextual information and marginalizes populations unrepresented by technology-derived data sets	Lacks scalability and statistical generalizability while being time and resource intensive	Possibly resource intensive, requires addressing problems of incommensurability

In applying big data within invigorated modes of urban research, then, there is a risk of over-simplifying some of the nuanced and complex realities of urban life and transformation. There is also a risk that simplified analytical understandings may underpin inappropriate or inadequate programs of city governance and development. For example, Datta (2015) argues that the concepts and manifestations of the "smart city" fail to connect with the plural realities characteristic of contemporary urbanisms in postcolonial India. There the smart city model all too easily overlooks the negotiations of citizens in their everyday encounters with the state as they seek to maintain their livelihoods while competing against top-down global aspirations of urbanism. "Smart solutions" that lean on big data techniques, when applied to issues such as urban water supply, can in practice "narrow down" complex problems into technical terms that privilege certain scientific knowledge claims over others while reframing citizenship as a mere practice of consumption (Taylor & Richter, 2017). For Hollands (2008), smart city approaches risk masking the real problems of the city and its past encounters with technological developments that have failed to address socioeconomic inequalities.

For Vanolo (2014), big data is a powerful techno-centric disciplinary tool that can be used to shape and govern the conduct of technologically literate citizens. Indeed, arguably one of the major limitations of big data analytics is the tendency to overlook the places and activities that fall within the shadows of data sets. The applications of such techniques may be limited to the most technologically privileged of society, addressing the needs and demands of more visible and powerful users and spaces while bypassing those of the less empowered (Graham, 2000). From a research perspective, then, data collected from mobile sources, online businesses (including social media), smart phone applications, and government portals can include various biases in the population samples employed, leading to inaccurate generalizations and conclusions (Arribas-Bel, 2014). Such biases can have particularly profound implications for studies undertaken in contexts of the Global South, where technological coverage may be less comprehensive and transient urban activities (such as informalized governance, mobility, or economic processes) can be difficult to capture through digital means. Aside from problems of examining and accounting for diverse urban realities, then, from a political perspective there is the potential for uncritical applications of big data technologies, under a guise of technical neutrality, to condition urban governance approaches that reproduce social inequalities and biases (Bannister & O'Sullivan, 2021; Benjamin, 2019; Taylor, 2021; Williams et al., 2020).

Recognizing some of these limitations, urban researchers working with big data sources and methods have explored opportunities for their combination with, for example, questionnaires (Sun et al., 2020) and field observations (Gabrys et al., 2016; Guan et al., 2021). However, to date there has not been any specific reflection on the value of ethnographic methods and knowledge for the realization of smart cities and emerging state-of-the-art urban analytics. While authors like Bibri and Krogstie (2017) uphold the value of advanced computational methods and big data for orchestrating smart and sustainable cities, they stop short of acknowledging the relevance and value of ethnographic and other qualitative methods for generating useful complementary knowledge of urban societies and processes. Thus, a kind of normalization of knowledge production—in the form of a mainstreaming of computational and algorithmic analysis—exists in how the smart or experimental city is imagined to function.

The value and limitations of thick data

By invoking the notion of thick data we refer to the kind of descriptive qualitative data collected and analyzed in the tradition of Clifford Geertz (1973), a cultural anthropologist who emphasized the importance of observing human behavior in the context in which it occurs in order to develop a holistic view of a phenomenon or practice in its wider social and symbolic dimensions. Thick data is descriptively rich and intensively detailed, collected from or with research participants on issues relating to their values, visions, knowledge, life experiences, and opinions, typically obtained by observing or interacting with participants in their daily lives and through in-depth interview techniques. Thick data may be collected by using and analyzing a variety of formats including photographs, videos, sound recordings, sketches, and written ethnographic field notes. The strength of thick data, by contrast to that of big data, lies in its attention to details and of objects, movements, practices, human behaviors, and words, for example. They can help to craft an in-depth and contextually nuanced understanding of peoples, cultures, and places that is a necessary condition for understanding complex urban phenomena.

Ethnographic approaches have important potential to contribute to our understanding of complex socio-cultural issues by examining them in their wider context and detailing the processes through which meaning is made. However, we should note various limitations relating to thick data and ethnographic methods, which big data and computational methods could potentially address. Ethnographic research is often resource and time intensive. Developing the linguistic, cultural, interpersonal, and local competencies required for ethnographic fieldwork calls for significant investment in time and resources. The length of fieldwork and data collection can vary from months to years. Creating and maintaining relationships of trust with communities and individuals is a long-term and continuous process demanding complex ethical sensibilities. These methodological requirements place limitations on the scale of analysis and on the possibility of conducting multi-site field studies simultaneously. Aside from these practical issues, at a more conceptual level ethnographers often face challenges in defining the spatial and temporal boundaries of what they are studying, and in determining the context that defines their object of analysis (Hammersley, 2006).

More broadly, within interdisciplinary conversations a number of critical questions have been posed regarding the philosophy of ethnography. Issues of generalizability, reliability, and validity remain key points of debate with respect to the strengths and limitations of ethnographic methods (Bernard, 2002; Hammersley, 1992; Schensul & LeCompte, 2012). Given their small-scale spatial and temporal purview and their intensive (rather than extensive) research focus, ethnographers are typically not able to generalize empirically from their findings to larger populations of cases across space and time. For these reasons, there are real advantages to tracing social behaviors in real time across a breadth of geographies using sensors and computational techniques (Pretnar & Podjed, 2019). That said, ethnography can produce grounded knowledge of patterns in cultural behavior and conduct while illuminating more general aspects of the human condition. Ethnographic knowledge is generalizable insofar as it seeks to situate the minutiae of urban life within their wider context, which can then be understood by readers through analogical forms of reasoning—the process by which people infer similarities or differences between the case under discussion and their own prior knowledge and experiences (Barnett, 2020).

Urban ethnography is by no means a new or novel approach, and there are numerous examples of empirical studies that take an anthropological view of the city, often striving to present an "insider" point of view in an effort to banalize or "make ordinary" what would appear to be bizarre or impenetrable urban phenomena (Duneier et al., 2014). Ethnography potentially allows for modes of theorizing the city that start from the particular and quotidian practices of the urban, casting in critical relief the universalizing pretensions of much urban theory emerging from the Global North (Lewis & Symons, 2017). For some, urban ethnography offers a unique contribution to our understanding of an increasingly complex, and increasingly urban, world (Pardo & Prato, 2018). However, to date this work has shown limited engagement with the sources and analytical methods associated with big data.

Emerging approaches to reconciling big and thick data

Few researchers and practitioners in any field have attempted to apply and reconcile big and thick data and their related methods of analysis. However, those who have done so offer a number of insights that are useful in considering how the limitations of emerging forms of big data–driven urban experimentation and governance can be counteracted. Writing from a sociological perspective, Bornakke and Due (2018) argue that big and thick data are compatible despite the limited attention that has been dedicated to their combination in research. They propose the term "blending" to describe an interdisciplinary process "in which insights based on big and thick data are brought together into new conceptualizations through deliberate actions performed by researchers" (p. 4). Their approach focuses specifically on bringing together different analytical insights, rather than different data collection methods, in "an interpretative, distributed cognitive and embodied process conducted by the researchers" (p. 4). Through a series of examples, the authors show how thick data collected through human observations allowed for a "calibration" of machine learning techniques and enriched the analysis of big data. Moreover, the use of big data enabled the phenomena identified through ethnographic observation to be productively examined at extensive scales.

Similarly, using observations and sensor-derived movement data of museum visitors, Girardin (2013) suggests that the articulation between qualitative insights and sensor measures enabled a more refined understanding of the phenomenon in question. Blok et al. (2017) recognize the granularity and density of big data in its extent of observations over space and time while zooming in on the details of digital social data, thus "stitching" together digital-transactional data and ethnographic observations to understand the dynamics of behavior at a student party in Denmark. They note the capacity of this kind of hybrid approach to integrate the spatially and temporally extensive qualities of big data with the "processual focus" of ethnography (p. 12). This enables inferences to be made about "collective life" at various spatio-temporal scales, in turn allowing "dynamic relations and patterns" to emerge from the data (p. 12). Meanwhile, in their analysis of the behaviors of faculty employees at a Slovenian university, Pretnar and Podjed (2019) refer to the iterative contribution of computational and ethnographic methods within "circular mixed methods" that enable researchers to study complex phenomena from different perspectives in alternating stages.

Other researchers have focused more on how the specific components associated with ethnographic research, such as the development of narratives and uncovering the symbolic nature of objects, can be brought to bear on data. Dourish and Gómez Cruz (2018), for example, empirically reflect on California's efforts to monitor paroled sex offenders. They draw on the practice of narration as a process of ethnographic tracing and the writing of "stories" in order to contextualize and bring meaning to data. They attend to the cultural embedding of data using a narrative framework to make sense of data within a "landscape of recognizable objects" (p. 6). Here the focus was on the role of objects within cultural settings and the "practices by which meanings are produced at particular times, in particular places" (p. 5). Thus, it is possible that data can become "symbolic" by taking on meaning, which in turn enables new kinds of "sense" to be made of those data.

Ethnographers have traditionally carried out their craft by studying situated, face-to-face, everyday interactions (Goffman, 1968). However, as online activity has increased, ethnographic approaches have also emerged for the examination of online social interactions. Social media data may now be "mined" in order to understand context-specific phenomena such as visual online culture, linguistics, and identity (Anderson et al., 2009; Magro, 2018). The emergence of the term "big data augmented ethnography" (Laaksonen et al., 2017) is evidence of the growing inclination for ethnographers to work with big data through online platforms in ways that bring new meanings to the situated encounters that are traditionally encountered in "the field." For example, when studying political discussions on social media, Laaksonen et al. (2017) posit that ethnographic observations can be used to contextualize the computational analysis of large data sets, which in turn can be applied to validate and generalize the findings made through ethnographic inquiry.

In a similar vein, Ford (2014) argues that ethnographers can provide "unique insight into how participants interact in complex media platforms." In their study on Wikipedia citation sourcing, Ford found that one specific observational detail gave insight into the social meanings attributed to citation sourcing practices; big data analysts were then able to test her hypothesis at scale. Meanwhile, Boy et al. (2016) use social media data from Instagram and big data-driven analytic techniques to identify spaces of cultural meaning in the city as well as to map and understand urban social divisions. They argue that urban researchers can mobilize the benefits of both quantitative and qualitative approaches to "investigate at a very large scale and in minute detail how urban dwellers form groups within and through urban space" (p. 13). In so doing, researchers can engage in iterative and collaborative ways of working to "develop different lenses" for analyzing big data.

Blok et al. (2017) refer to these kinds of processes as the "contextualization" of big data. The studies discussed here offer new hybridized ways of collecting and analyzing thick and big data; approaches that fuse methods and insights from both domains. In these studies we see the blurring of epistemological boundaries as well as the roles afforded to the categories of "big data scientist" and "ethnographer." For example, Curran (2013) offers the term *Big Ethnographic Data* while arguing that ethnographers should see their task as being more than producing insights on the "hows" and "whys" that big data research needs, rather viewing big data and ethnography as essentially the same. This could involve focusing on how big data is used as a process to interpret culture, and how big data in turn shapes culture.

To overcome some of the limitations of understanding the city using only a big data lens, there is a growing need for the application of more insightful data that can help to refine, calibrate, contextualize, or interpret routinely collected big data (Wang, 2013). At the same time, we need to recognize the capacity of big data analytics to scale, extend, validate, and generalize the insights derived from thick data. Therefore, in order to contribute to emerging fields of scholarship and practice related to city science and smart cities, we consider how ethnographic practices can be combined with big data analytics in the examination of critical urban problems. We see this as part of wider effort to move beyond the hostile qualitative-quantitative dichotomy that remains deeply embedded in the disciplinary nature of academic research, and that arguably has been exacerbated by the emergence and influence of experimental and data-led urbanisms. We also see this as a way of expanding encounters between ethnography and big data beyond online space and the domain of digital culture.

Why should such a reconciliatory approach be particularly suited to or needed for the analysis of *urban* processes? One answer is that cities, as inherently complex objects, demand an interdisciplinary research approach (Mora et al., 2021; Verloo & Bertolini, 2020). Moreover, a view of cities as complex open systems (Amin, 2020; Bai et al., 2016; Barnett & Parnell, 2016)—enshrined in the post-2015 global urban agenda for sustainable development—enjoins a methodological approach capable of understanding the dynamics of urban change at multiple temporal and spatial scales (Creutzig et al., 2019; Duminy & Parnell, 2020; Smith et al., 2021; Elmqvist et al., 2018; Frantzeskaki et al., 2021). There is no reason why such approaches would not be suitable for examining less urban or rural contexts, but the demand for methodologies that traverse disciplines, systems, and scales in search of "leverage points" (Leventon et al., 2021) for transformative change is particularly prominent in the case of urban sustainability research. A reconciliation of big and thick data can provide a perspective necessary to grasp the dynamic and multi-scalar nature of urban emergence and experience (Keith et al., 2020). We see this discussion as a complement and contribution to a growing chorus of calls for a more systematic and integrated "science of cities" (Batty, 2012, 2013; McPhearson et al., 2016)—one that is capable of grasping the global urban condition at scale while accounting for factors of specificity and difference (Parnell & Robinson, 2017).

Our reflections, described in the following sections, move beyond more established and conventional approaches to mixed-methods research, in which important conversations have been had surrounding the application of digital technologies albeit without a specific interest in the gathering and application of thick data (O'Halloran et al., 2018, 2019). We specifically consider how big data analytics and ethnography could be combined within a more adaptive and flexible approach to urban inquiry. We envisage that such an approach will be relevant to situations in which it is not possible to use only secondary data, or to collect data systematically using digital methods, as is often the case in developing countries or contexts hosting marginal, vulnerable, or displaced populations.

An illustrative example of a reconciled approach: Road safety in Mexico City

Through our group discussions, involving not only urban big data scientists and mathematicians but also ethnographers and social scientists from various disciplines, we constructively considered how to go about designing and collating various methods to produce the data and knowledge that would enable us to understand and respond to a key urban problem: road injuries and fatalities.

The contributing authors are all participants in the PEAK Urban research project funded by the UKRI's Global Challenges Research Fund. The objective of the PEAK Urban project is to bring together early-career urban researchers from five countries (United Kingdom, India, China, Colombia and South Africa) to develop innovative and interdisciplinary ways of dealing with complex urban problems (Keith et al., 2020). As part of the project, three "retreats" were planned for the participating researchers. Each retreat involves a series of workshops, discussions, presentations, research engagements, and writing exercises to help develop interdisciplinary modes of working. The contributing authors are one of several interdisciplinary "working groups" that have developed around specific self-selected research themes. Our common interest in developing new ways of integrating data and in overcoming historical, ontological, disciplinary, and methodological boundaries and conflicts brought us together. Since we formed in early 2019, following our first retreat held in Bangalore (India), we have held seven online group discussions. Three of us are scholars who work with big data methods, while four collect and use thick data on studies covering urban mobility, informal housing and migration, health and wellbeing, and urban governance. The discussions were an opportunity to learn about different approaches that could be used to examine urban problems such as road safety, the insights that could be gained from adopting various approaches, how they might complement one another, and how relevant changes in policy might be developed.

Applying a big-thick hybrid approach to urban mobility analysis and planning

Our group selected road safety as an illustrative example of how big and thick data and related methods could be reconciled in practice. It is a relevant example in part because the field of transportation has a history of adopting technical-rational approaches that have often failed to understand urban mobility empirically, in context, and in relation to the historical pathways through which social problems and their solutions, along with their complex structures of power, emerge (Marsden & Reardon, 2017). That said, researchers active in the fields of transport and mobility studies increasingly recognize the importance of culture and place in understanding social phenomena and are turning to qualitative approaches to understand complex problems in a more holistic way, and as part of what Schwanen (2017) describes as the reconciliation of "the general" with "the particular."

Big data analytics has made important contributions to our understanding of urban mobility patterns (Noulas et al., 2012; Shaw et al., 2016). For instance, by capturing credit card transactions and purchases, researchers have been able to map out urban lifestyles (Di Clemente et al., 2018). Mobile phone data have allowed the detection of distributions of travel demand within cities (Widhalm et al., 2015) as well as daily patterns of mobility (Schneider et al., 2013). Mobility data allow us to investigate complex phenomena relating to human movement (Rinzivillo et al., 2014), to classify different mobility behaviors (Pappalardo et al., 2015b), and to link the dynamics of mobility with levels and distributions of income and wealth (Pappalardo et al., 2015a). Recent research has taken advantage of big data obtained from a range of sources (Chen et al., 2018; Xie et al., 2017) including in-vehicle devices (Azmin et al., 2018) and social media (Chen et al., 2017) to analyze the dynamics of driver behavior and traffic safety. Some of this work is motivated by demands from the insurance industry to charge premiums based on individual driving behaviors (Arumugam & Bhargavi, 2019).

Scholars influenced by the "mobility turn" in the social sciences led by Sheller and Urry (2006), Adey (2006), and Cresswell (2006), among others, have turned to qualitative methods to understand mobility patterns and dynamics in greater depth than mainstream quantitative approaches previously allowed. For example, scholars have explored complex issues of power and injustice in the field of mobility (Jensen, 2009). Some have examined how certain subjects including women, minority ethnic populations, the elderly, and disabled persons experience lower levels of mobility and accessibility in urban areas, often disadvantaged by the ways in which cities have historically been planned (Urry, 2016). Others, influenced by anthropologists, have adopted ethnographic approaches to observe and uncover the highly contextualized and complex nature of urban mobility patterns (Jensen, 2009).

Inspired by these emerging bodies of scholarship, we set out to consider how a variety of thick data methods and insights might be used to complement big data practices in researching issues of urban mobility. In order to demonstrate what we have in mind, we offer the illustrative example of road safety issues in Mexico City. One of our group members hails from Mexico City, and previously he has actively participated in public discussions surrounding road safety and management issues. We considered that it provided an ideal example through which to orient our group discussions and reflections.

Addressing road safety in Mexico City

According to the World Health Organization, in 2016 the global number of annual road traffic deaths reached 1.35 million, and globally between 20 to 50 million people suffered a severe injury, many resulting in a disability (World Health Organization, 2018). Comparatively speaking, violence—

including all types of crime, war, and conflict—causes fewer fatalities (estimated at 560,000 annually). Globally, road traffic injuries are the leading cause of death for people aged between five and 29 years, but this problem is more acute in low and middle-income countries, in which more than 90% of road fatalities happen. In recognition of the scale of the problem, the General Assembly of the United Nations proclaimed 2021–2030 as the Decade of Action for Road Safety. Moreover, reducing injuries as a result of road accidents is emphasized by the Sustainable Development Goals; target 3.6 aims to halve the number of global deaths and injuries from road traffic accidents by 2020. Unfortunately, that target was recently revised with to a 2030 timeframe after the 2020 deadline was missed (Mohan et al., 2021). Nonetheless, increasing road safety remains a global priority for urban governance and intervention.

Causes of road injuries and fatalities are many, but according to the World Health Organization key contributing factors include speeding (Elvik et al., 2004), driving under the influence of alcohol and other substances, being distracted (Lyon et al., 2021), driving while exhausted (Williamson et al., 2011), poor infrastructure, unsafe vehicles, nonuse of safety devices (such as helmets or seat belts), inadequate enforcement of traffic laws, inadequate post-accident care, as well as other environmental factors including weather (Edwards, 1996).[1]

Mexico City has a high prevalance of road traffic injuries and fatalities. From 2015 to 2019, the city experienced major policy reforms relating to road safety (described below) including initiatives to include an examination as part of the process of obtaining a driver's license (this is still not in place[2]) and the development of automated ticketing systems. By leveraging recently published open access data on road accidents in Mexico City,[3] we consider the example of road safety to illustrate the complexity of urban problems and to demonstrate how such issues can be productively explored using sources and techniques of big and thick data.

Road injury and fatality data on Mexico City reveal a clear spatial pattern (Figure 1) with a high degree of temporal and spatial concentration (Prieto Curiel et al., 2018, 2021). Main avenues and highways, as well as specific junctions, are frequently highlighted in our map (Thomas, 1996). The map also reveals clusters of accidents around some Metro stations and other public transport nodes with high pedestrian flows, particularly in the city center (Prieto Curiel et al., 2021). Road accidents are highly concentrated around zone junctions and in motorways that originate in the city (Prieto Curiel et al., 2018).

Although some spatial patterns are evident in the map, and various temporal patterns are also evident through data visualization techniques (Prieto Curiel et al., 2021), it is not possible to detect the specific causes of road injuries and fatalities, nor potential policies or interventions that could assist in their reduction. Further inquiry into the data suggests that the majority of road traffic injuries and fatalities (82%) involve two or more motor vehicles, whereas fewer incidents involve cars and pedestrians (15%) or motorcycles (3%). Here we will focus specifically on road fatalities and injuries involving motor vehicles only, which requires a specific understanding of driver behaviors and attitudes to develop appropriate policies and interventions to reduce traffic incidents. We limit our discussions to Mexico City, as other cities may have different mobility and road safety patterns involving pedestrians, cyclists, or motorcycles that require an understanding of additional or alternative contextual factors at work.

A big data approach to road safety

The data set used as a starting point for our approach consists of approximately 1,100,000 road accidents, including minor events, that have taken place since 2015 in the central area of the Mexico City metropolitan area (hosting 8.8 million inhabitants, with an estimated 22 million living in the wider metropolitan area).[4] The data include the spatial coordinates, time of day, date, and number of casualties for each event, among other factors. Although under-reported traffic injuries are rare, particularly if they are severe, there is still a high chance that minor road incidents are not captured by this data (Savolainen et al., 2011). As such, the data neither are of the scale to be considered "big" nor are they collected in real-time. Nonetheless, the collection and processing of this data does require

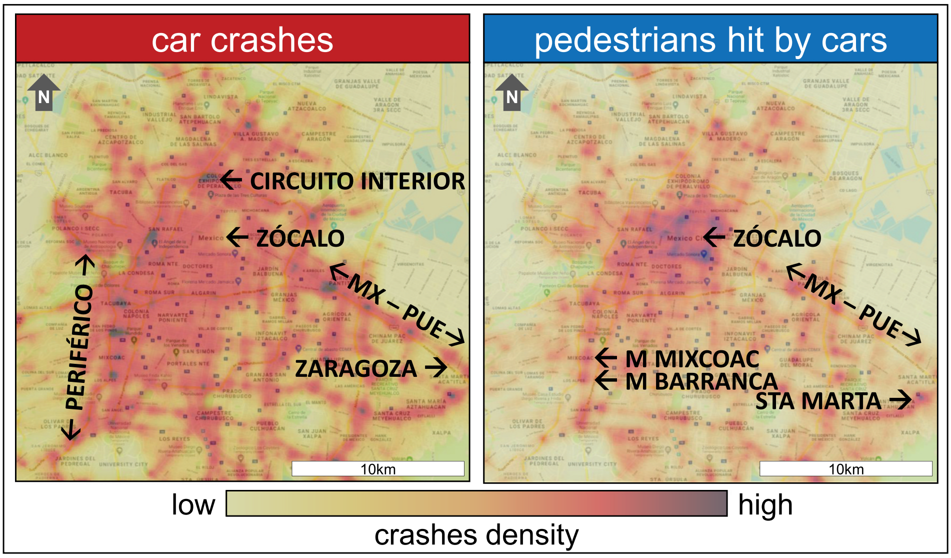

Figure 1. Heat map of serious and severe road incidents in the central part of Mexico City (not the metropolitan area), created using Open Access data from 2015 to May 2019. The left panel shows crashes involving only cars and the right panel shows crashes involving cyclists and pedestrians. Main avenues, such as Anillo Periférico, Circuito Interior, and Calzada Ignacio Zaragoza and certain junctions, such as Insurgentes and Circuito Interior, are the main road traffic injury and fatality hotspots in the city, as well as the areas nearby some major Metro stations, such as Metro Barranca, Mixcoac and Santa Martha.

a certain level of computational power. Moreover, this data set provides a useful starting point for a discussion regarding the development of an interdisciplinary research approach that draws upon big and thick data to understand the problem of urban road traffic fatalities and injuries.

Previous studies of road accidents show that the availability of geo-located microdata allows for the analysis and understanding of spatial patterns of traffic safety and can help urban planners and transport engineers to identify areas in which incidents are more likely to occur (Erdogan et al., 2008; Fan et al., 2019; Prasannakumar et al., 2011). Thus, traffic microdata have been used for identifying hazardous road segments (Bíl et al., 2013), regions in which pedestrian injuries are common (Schuurman et al., 2009), in addition to other relevant factors. Using this type of data, it is possible to construct a spatial heat map showing traffic safety hotspots (Anderson, 2009), which tend to correlate with traffic volumes (Thomas, 1996) as well as other socioeconomic and environmental factors (Aguero-Valverde & Jovanis, 2006; Steenberghen et al., 2010).

There are a number of ways to further refine these insights, however. For example, the results can be normalized by traffic flow density and built environment intensity to reflect the true hotspots of traffic safety. Moreover, additional spatial analysis and synthesis can be performed. The temporal dimension can also be applied to represent time series of spatial alternations. Big data could be used to identify locations in which more accidents occur due to higher traffic volumes at certain times of the day (Prieto Curiel et al., 2021). It may also identify poorly designed road infrastructure (Ge & Fukuda, 2016) as measured by, for example, the density of intersection, the redundancy of local streets, and the width or number of lanes of thoroughfares. In addition, geo-referenced social media data can be used to detect distracted driving behaviors (Doran et al., 2016), and related data mining techniques can be employed to develop profiles of traffic incident hotspots (Kumar & Toshniwal, 2016; Kwon et al., 2015). In these ways, big data techniques can be used to quantify and spatially map traffic fatalities and injuries to provide important insights enabling the development of appropriate policies to improve traffic safety (Stehle & Kitchin, 2020).

While large-scale traffic accident data can help identify important spatial patterns at the macro level, other types of big data technology and techniques can be used to capture individual-level driving behaviors linked to traffic safety. Technologies that monitor driving behaviors have existed for some time, but with recent advances in artificial intelligence and computing power, combined with the falling cost of sensors, it is now possible to collect and process large amounts of driving behavior data to uncover previously hidden details about individual driving patterns in real-world situations. For example, a prominent approach applied in the field of traffic behavior research is the naturalistic observation of driver behavior (Klauer et al., 2006). Under this approach, details of the driver, the vehicle, and the surrounding environments are recorded and stored using unobtrusive sensor technologies, such as in-vehicle radar, video cameras, accelerometers, Global Positioning System (GPS), and on-board computer vision algorithms.

In our example of Mexico City, big data generated from these sensor devices could be used to trace the impact of unusual driving patterns on road accidents. Drivers exceeding speed limits could be tracked with the use of real-time GPS data (Li et al., 2020). A combination of GPS, accelerometer, smartphone data, CCTV data, and sensor data can be used to detect anomalous behaviors such as driver inattention (Lyon et al., 2021), distraction, or sleepiness (Kanarachos et al., 2018). These sensing technologies allow researchers to not only collect a wealth of data regarding normal driving behavior, but also capture data on any crash or near-crash events as the data are being recorded continuously over an extended period of time, usually over a year. This extends the ability of big data to answer questions regarding the behavioral symptoms or possible causes of traffic accidents. Indeed, one of the first comprehensive studies to use the naturalistic observation of driver behavior reported that 78% of all crashes and 65% of near-crashes were attributable to driver inattention (Klauer et al., 2006). In Mexico City, similar naturalistic observation techniques could be adopted to provide a deeper understanding of the processes resulting in crashes and near misses, allowing researchers and law enforcement officials to identify the causes and conditions of safety-critical events.

A thick data approach to road safety

Many aspects of road injuries and fatalities would not be captured through the techniques discussed above. In particular, important gaps would remain with respect to the attitudes of drivers and the political and societal impetus for action in the domain of traffic safety regulation. Moreover, big data analyses may overlook the historical pathways giving rise to driver attitudes; pathways that are embedded in specific contexts and that over time lead to the emergence of established social norms. For example, in a recent study interviews with drivers in Mexico City revealed that young drivers tended to experience higher levels of anger and were more likely to express that anger by comparison with more mature drivers (Hernández-Hernández et al., 2019). Given the significance of such factors in determining unsafe driver behavior, be it risky, aggressive or problematic in other ways, it would be necessary to consider the complex realities of driver attitudes with respect to road injuries and fatalities prior to developing any policy recommendations based on big data analytics. Below, we outline three factors influencing the ecology of road safety and driver behaviors, our understanding of which would be enhanced by ethnographic approaches, methods, and “thick” insights.

The political nature of law enforcement

Prior to 2015, there was little systematic enforcement of traffic speed limits in Mexico City. However, at this time a new automated ticketing system was introduced with the objective of punishing drivers who violated speed limits and other regulations. The automated system increased awareness of driving risks while deterring drivers from speeding via a monetary penalty. In 2018, however, a newly elected government decided to retract the fining system with the argument that it was exerting a detrimental economic impact on Mexican families. Contrary to this claim, data showed that the frequency of drivers penalized with a speeding fine increased with the value of an offender’s vehicle. As a result, the people most likely to be affected by the stricter enforcement of speeding limits were those who were least likely to suffer detrimental economic impacts from the fines. If the value of a vehicle is attributed as a proxy for socioeconomic profiling, then the data suggest that as an individual's economic status increases, their propensity to violate traffic speed regulations also increases.[5]

As noted above, Mexico City underwent major elections in 2018 before the ticketing system was removed. A number of political candidates ran on a vehicle-centric manifesto to gain support by promising the removal of the system. The manifesto was justified by arguments pointing to the system's excessive financial costs and poor design. The enforcement of traffic speeding regulations in Mexico City was exploited by politicians to foster political patronage in the buildup to the elections. The example demonstrates that road safety is not a politically neutral urban problem. Rather, it is a problem reproduced by and through political actors and interests. Without the political will and consistency to implement measures targeting incident hotspots identified through big data techniques, the application of such techniques will be undermined and potentially blocked. Consequently, developing a refined understanding of a city's political and economic context—including the historical emergence of political networks and interest groups, and their relations with particular kinds of urban infrastructure elements—would be essential to the embedding of big data-driven methods within city governance.

Ultimately, between 2019 and mid-2021 the daily number of traffic-related casualties in Mexico City increased by 4.7%.[6] It is not possible to identify a single cause for this increase, which occurred in a period of significantly reduced mobility linked to the COVID-19 pandemic and related restrictions and stay-at-home orders. Still, such a growth in the rate of lethal traffic incidents in the city has not been observed for at least a decade, suggesting that the enforcement of traffic speeding regulations is not effective in preventing casualties.

Driver attitudes and behaviors

The attitudes of individuals with respect to road incidents can in part be a response to the accidental nature of many road incidents, as such it may be perceived that little can be done by way of prevention (Lamont, 2012). In the context of Mexico City, the authors' personal observations suggest that it is not rare to find drivers who think that they "can drive better" when they are under the influence of alcohol or traveling at high speeds. Indeed, a dominant mentality exists among many drivers that "those who crash are those who cannot drive" (DeJoy, 1989; Delhomme, 1991). Both in-depth interviews and field observations can be used to understand risky driving behaviors and attitudes toward driving that are shaped by wider social norms (Sanusi & Emmelin, 2015). Such an approach might interrogate social attitudes related to drunk driving, for example (Horwood & Fergusson, 2000). It could do this indirectly by, firstly, understanding cultures of alcohol consumption in greater detail through observations and, secondly, relating these cultures to the use of vehicles and patterns of mobility. For example, conversations with drivers might reveal gendered ways in which driving fast is normalized, and how these norms in turn embody and express identities that are intensified under the influence of alcohol or in other kinds of social situations. Attitudes may change depending on a driver's age, reflecting not only their level of experience but also the different ways in which identities are embodied through driving practices among particular demographic cohorts (Falk & Montgomery, 2007; Ramos et al., 2008; Watters & Beck, 2016). Alternatively, conversations may reveal that a lack of travel options exists in spaces in which people gather to drink alcohol socially at night, which steers them toward the risks entailed with drunk driving. Observations of drivers who are likely to cause traffic incidents (identified by quantitative analysis) may generate a deeper understanding of driver identity, behavior, and perceptions of risk. This would enable further exploration of when risky behaviors are most likely to occur, thereby linking driving patterns with the rhythms of everyday life, emotions, and the availability and quality of infrastructure in different city spaces (Moran et al., 2010; Pilkington et al., 2014).

Social and spatial divides

The heat map presented above offers information about where traffic incidents occur more or less frequently. However, it provides little knowledge of why this spatial pattern has emerged. Road safety research has highlighted the effects that infrastructure and the wider socioeconomic and built environment have on the occurrence of traffic injuries (Cabrera-Arnau et al., 2020; Najaf et al., 2018; Schuurman et al., 2009). Ethnographic research might therefore explore the fabric and rhythms of the spaces identified as having higher or lower occurrences and reports of incidents. Through observations in the field, one might ask what it is about a certain space that reduces, or increases, road safety hazards. An ethnographer could detail the spatial patterns of land-use or the social relationships, behaviors, and recognizable social codes that are shared and normalized within specific road spaces to give rise to injuries and fatalities (Holmes et al., 2019). For example, some areas of Mexico City have higher concentrations of bars and might thus embody a culture of nighttime socializing centered on the consumption of alcohol. Observations could determine how factors relating to the spaces in which incidents occur, including the social norms and environmental factors active in those spaces, interact to produce certain kinds of risks at different times. Furthermore, in situations where big data is generated through technological devices, there is a risk of data gaps emerging, particularly around informal activities of urban settlement, economic exchange, service provision, and governance. Given this risk, an ethnographic approach could be used to trace appropriate local stakeholders and to access communities known to be experience barriers to using digital devices contributing to big data analytics. These communities might assist, for example, by participating in mapping exercises and go-along methods, to understand their experiences and to identify accident hotspots that are undetectable by big data methods. Thick data might thus illuminate risk-generating patterns of land-use and inappropriate infrastructure, the spatial politics of where safer infrastructures are implemented, the ways in which infrastructures are used, as well as a range of other environmental factors that contribute to the occurrence of road incidents (Drottenborg, 1999).

Discussion: Identifying opportunities to reconcile big and thick data, and insights for a hybrid approach

In the previous two sections we considered ways of examining and addressing road safety problems using thick data and big data separately. Here we critically discuss how big and thick data approaches might come together to develop, strengthen, or expand on the above methods and suggestions. We anchor our reflections around three key phases of the research process, namely (1) the formulation of research questions, research design, and stakeholder engagement; (2) data collection and analysis; and (3) knowledge representation and the development of research outputs. In setting out our discussion in this way we do not seek to present these phases as discrete and non-overlapping, to be undertaken in a linear manner. Rather, we believe that one of most significant contributions of taking a more integrated approach, involving both big and thick forms of data, is the capacity to generate new insights and approaches that extend across all stages of the research process. This iterative process is illustrated in Figure 2.

Formulating research questions, research design, and stakeholder engagement

Undertaking urban research in ways that incorporate the insights of both big data analytics and ethnographic description can help us to recognize the political and cultural dimensions of what may appear to be purely technical problems. Problems of urban development such as road safety seldom have simple causes and solutions, but rather activate a wide range of actors and processes and therefore demand a multifaceted research response. The gathering of thick data can help to unsettle the understandings and assumptions on which research questions and designs are based. It can remind us that research questions are simply a starting point requiring constant reflection as new data and information emerge through the research process.

Second, ethnographic research will bear important insights to assist with the identification of relevant constituencies and to help determine how they should be engaged in the research process. For example, informal street or neighborhood-level structures may exist and have an important influence on how local road safety interventions are received and implemented in

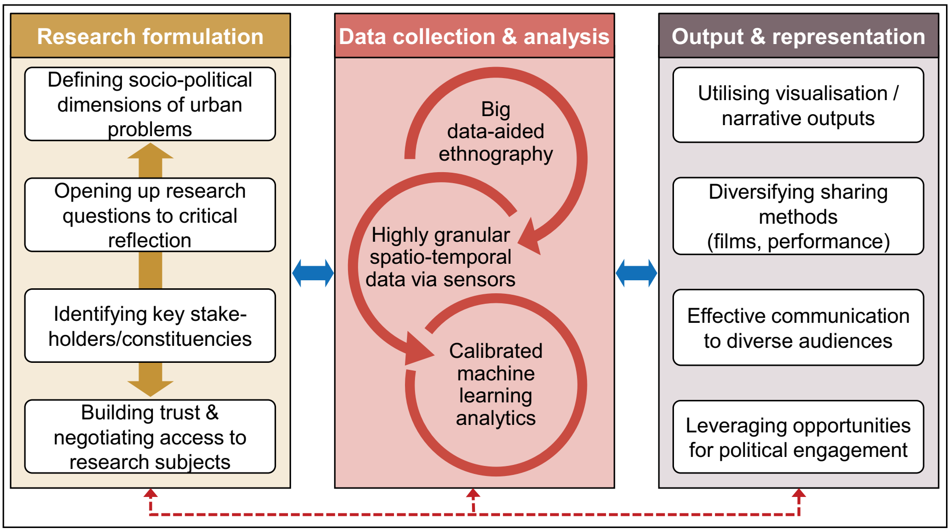

Figure 2. An iterative process of synthesizing big and thick data across three stages of research.

areas of Mexico City, yet these structures might not be readily visible to researchers working exclusively with big data. In the course of gathering thick data at the local level a number of different interest groups would likely be identified. For instance, big data can be used to identify subgroups of drivers to be targeted for further ethnographic observation and through which specific behaviors can be examined in detail. Such a degree of focus may then reveal driver profiles that were not previously apparent or obvious, and that may be more or less specific to a particular city. Conversely, thick data may calibrate the use of big data, for example, by observing cultural behaviors in bars or on social media platforms to identify at-risk subjects who could then be traced at a broader scale using more extensive analytic techniques. In this way, the insights born of big and thick data analyses would complement and feed into one another as part of a larger iterative and circular process of research design.

Third, ethnographic researchers are required to build sustained engagements and trust with their research subjects, and consequently will often develop an acute understanding of the particular sensitivities involved in approaching and securing the consent of community groups and other stakeholders. That kind of understanding is critical for urban researchers to engage with local constituencies in a sustained and effective manner. Big and thick analytical modes of inquiry could be carried through an overall process of place-based research coproduction through which data sources and research needs are determined collectively by the full range of stakeholders affected by a particular problem, and whereby both big and thick data feed into a common and local understanding of the dimensions and drivers of that problem. This could be of particular advantage in encouraging approaches to data-driven research that incorporate the logic and procedures of citizen science—approaches that would mobilize beneficiaries themselves as gatherers and owners of data as part of a larger democratization of urban knowledge production (Gabrys et al., 2016). Such approaches may help to overcome resistance from potential research subjects arising from their individual concerns around digital privacy and surveillance and the possible effects of increased surveillance for societies at large. They would also assist with the contextualization of big data at the design phase, refining its application to be relevant to specific places and to ground the meaning of its insights with the needs identified by citizens.

Data collection and analysis

While ethnographers can generate important insights into the details of urban problems and their causes, the analysis of big data has a crucial role to play in making wider statements and generalizations about the prevalence of certain kinds of behaviors and processes across scales. This might include, for example, the extent of certain kinds of risky driver behavior and their distribution across urban populations in space and time. At the same time, that analysis would benefit from field observations through ride-alongs or interviews with drivers in order to identify specific aspects of risky behavior that can be used to inform the monitoring of certain spaces and groups. Ethnographic insights would further help to calibrate machine learning algorithms to detect risky driving behaviors in real-time to trigger one or more responses to prevent incidents from occurring. Moreover, ethnographic analysis applied to social media posts and online data, sometimes referred to as "Netnography" (Kozinets, 2015), could be used to better understand driver behavior and the construction of spaces of risk in a context like Mexico City. Having a refined sense of how traffic spaces are constructed as cultural entities characterized as risky, speedy, or dangerous would be an important complement to the predictive and correlational power of big data analytics.

The temporal granularity afforded by big data potentially enables city managers and planners to undertake real-time adjustments of driver behaviors to prevent traffic risks. However, while one can envisage a traffic management system that automatically detects erratic driver behavior and deploys policing responses, ultimately the specific kinds of measures and interventions that will be considered acceptable will vary according to social and political context. Gathering thick data would be particularly relevant for developing an understanding of the specific concerns held by various stakeholders, and how these might be overcome. This point links to issues of knowledge representation, dissemination, and implementation.

Research outputs and knowledge representation

Combining big data analytics and ethnographic research may encourage the generation of new modes of representing knowledge and novel kinds of outputs that communicate complex urban realities in ways that further our understanding of urban problems and facilitate the development of effective responses.

A principal benefit of data-driven urban analytics is the capacity to generate visualizations of research findings that effectively communicate complex and extensive urban realities in forms that are accessible and understandable to most actors and beneficiaries. Likewise, the detailed narrative form of case-based knowledge arising from ethnographic work has particular advantages for learning and helping actors to develop their capacity for situated ethical judgment (Flyvbjerg, 2004). Combining the two within discrete research outputs might be particularly beneficial. There are already important examples of how this might be done, including through the development of "data stories" (Gabrys et al., 2016) and "story maps"; resources that enable users to combine the information generated through mapping with the influential and awareness-creating power of storytelling within a single frame.

A particular issue facing researchers working co-productively in urban settings of the global South concerns how one should communicate results with targeted users and groups that may not be digitally or scientifically literate. In such circumstances, the use of narrative devices and stories may take on even greater importance, and the use of films, performances, photography, and other methods can offer important devices through which to share knowledge, secure feedback, and develop trust and mutual understanding. Some urban disciplines have been more innovative and adept at these issues than others. Health researchers, in particular, have long recognized the importance of research communication (Stewart, 1995) and the use of innovative approaches to implement findings among target users, including through street theater, radio or television programs, gamification, and so on. We would argue that researchers working on other pressing urban problems could do similarly, and that in doing so they would benefit greatly from the collection and analysis of both big and thick data.

Ultimately, the potential of big data to underpin new and effective modes of urban governance and planning will depend on issues of political will and acceptability. This is another area where thick data can make a key contribution to data-driven analytics, by helping to develop a nuanced sense of the socio-political context in which certain kinds of urban responses become more or less feasible, and the particular kinds of messaging that should be carried out in order to promote uptake and implementation across a range of actors and constituencies.

Conclusion

Big data technologies are a key foundation of new and experimental modes of urban research and governance represented by the notions of smart cities and the new urban science. However, concepts and practices related to big data may engender a number of methodological concerns and limitations. The movement toward data-driven urban analysis has followed a trajectory that diverges from the strengths of thick data and the insights derived from ethnographic approaches to seeing and knowing the city. The divergence of thick and big data in emerging ways of understanding the city is concerning and likely to limit, rather than enhance, the efficacy of the processes through which we respond to pressing urban problems. With reference to an illustrative example of road safety in Mexico City, in this paper we have provided an initial framework for how big data and thick data could be reconciled within urban inquiry.

We have argued that reconciliation can become a successful interdisciplinary practice when different methodological approaches and insights are applied at three crucial phases of the research process. First, at the stage of formulating research questions, reconciliation would help us to conceptualize common objects of inquiry in all their interrelated dimensions. It allows us to recognize the political and cultural dimensions of urban problems that may be assumed to be primarily technical in nature. It has the potential to identify a wider range of research subjects and urban problems, studying these through a variety of methods that are each uniquely appropriate to particular situations and scales of observation. Moreover, big and thick data approaches would combine methods and activities with the potential to meaningfully engage research beneficiaries in knowledge production as part of an open and democratic research process.

Second, with respect to data collection and analysis, reconciliation would help us to diversify our data sources as well as promote interactive feedback between different methodological approaches. This would enable an ongoing refinement of the inquiry and the adaptation of data gathering approaches according to emerging insights born of the analysis. Understanding how urban spaces are constructed as cultural entities and characterized in particular ways can complement the real-time predictive and correlational power of big data analytics. In turn, reconciliation could facilitate conversations between research subjects and analysts around issues including concerns of privacy, security, and trust.

Third, in terms of research outputs, reconciliation would help us to tailor our knowledge products, in their forms and representations, to diverse stakeholders. Wider stakeholder engagement with diverse research outputs creates opportunities to forge institutional and actor-based knowledge partnerships that can then inform and shape future research inquiries.

The reconciliation of big and thick data, working at different sub-phases and micro-processes of each main stage of a research project, has the potential to generate the deep interdisciplinary knowledge required by urban specialists (Reese, 2014). Reconciliation can thus be viewed as a new interdisciplinary sensibility, which when applied with good-faith conversations will produce concrete research practices that complement and enhance the explanatory and predictive power of the urban sciences.

Urban researchers seeking to work in an interdisciplinary mode will always face profound difficulties in finding common vocabularies and in overcoming the incommensurability of different methods and bodies of knowledge. One of the key challenges in undertaking successful interdisciplinary dialogs remains our hardened situatedness among our respective disciplinary methods, conventions, convictions, and comforts. As urban researchers, many of us express a reflexive desire to hold interdisciplinary conversations and to find ways of collaborating on common projects, yet we often struggle to realize our aspirations through concrete plans and practices. Reconciling big and thick data will always be a tension-filled process, but such tensions can also be productive. For this reason, it is useful to think of the integration of big and thick data and their related techniques as a conciliatory process and an always-emerging work in progress. Yet we believe that such an approach is imperative if we are to mount an effective response to the urban challenges of our time.

Notes

1. From https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries
2. https://www.aljazeera.com/news/2018/01/mexico-city-reform-traffic-laws-accidents-18010912.html
3. https://datos.cdmx.gob.mx/pages/home/
4. Open Access data collected by the Emergency Attention Center of Mexico City has information with respect to road accidents in the city, available at https://datos.cdmx.gob.mx/pages/home/
5. From http://www.eluniversal.com.mx/autopistas/los-autos-con-mas-fotomultas-en-la-cdmx(in Spanish)
6. From https://www.semovi.cdmx.gob.mx/tramites-y-servicios/transparencia/reportes-e-informes/hechos-de-transito

Acknowledgments

This work was made possible thanks to the support of the PEAK Urban program, funded by UKRI's Global Challenge Research Fund, Grant Ref: ES/P011055/1.

Disclosure statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This work was supported by the UK Research and Innovation (UKRI) Global Challenges Research Fund [ES/P011055/1].

About the authors

Andy Hong is Assistant Professor in the Department of City & Metropolitan Planning at the University of Utah. He is an honorary research fellow at the University of Oxford and the George Institute for Global Health. Andy is also co-founder of the Healthy City Futures, a global nexus of innovators dedicated to sharing cutting-edge information on urban health. His research interests lie at the nexus of urban planning, transportation, and public health.

Rafael Prieto Curiel is a Postdoctoral Research Fellow at the University College London (UCL) Centre for Advanced Spatial Analysis (CASA), working on urban dynamics as part of the PEAK Urban project. He also works for the OECD, performing spatial and demographic analyses of African cities. He holds an MSc and PhD from UCL in mathematics and security and crime. He has worked in the Emergency Attention Centre in Mexico City (C5), where he was Director of Strategic Analysis. There, his work consisted of crime forecasting and police and resources allocation.

Lucy Baker is an honorary research associate at the Transport Studies Unit (TSU) at the University of Oxford. Prior to this she was a research associate in urban mobility within the PEAK project based at the TSU, where her research focused on big data, digital technologies and concepts of the "smart city," and their applicability to mobility governance in India.

James Duminy is Lecturer in Human Geography at the University of Bristol and Honorary Research Associate at the African Centre for Cities (ACC), University of Cape Town. His work examines the governance of urban change in the global south. As a postdoctoral researcher in the PEAK Urban program based at the ACC, he examined historical and contemporary processes of national urban reform in South Africa.

Bhawani Buswala is a postdoctoral researcher in the PEAK Urban and Informal Cities programmes based at the University of Oxford. His principal research interests include questions of social inequality, urbanization, migration, discrimination, and the state. He explores these themes through ethnographic data on informal settlements, spatial segregation, everyday lived experiences, and claims on the state. His fieldwork is situated in North India.

ChengHe Guan is Assistant Professor of Urban Science and Policy at NYU Shanghai. He serves as a research consultant at the Centre on Migration, Policy and Society, University of Oxford. Dr. Guan is the co-director of NYU Shanghai Urban Lab. His research interests include (1) computational urban science and space-time big data analytics, (2) planning policy-informed urban growth simulation using CA-based and machine learning algorithms, and (3) urban vital sign system and digital transformation of low-carbon cities. He has taught multiple courses including "Urbanization in China," "Planning Global Cities," "Spatial Dynamics of Urbanization," and "Computational Urban Science."

Divya Ravindranath is a faculty member and researcher at the Indian Institute for Human Settlements (IIHS), Bangalore. Her research examines how the nature and conditions of work affect health experiences and outcomes among those employed in the informal economy, in sectors such as construction and domestic work. In a recent project, she examined the impact of paid and unpaid work on breastfeeding practices at construction sites. This project also explored different methods used to study time-use. Her ongoing research work on social protection and health has a strong policy and practice-focus. In the past, Divya has worked with several not-for-profit organizations on issues of livelihoods, infrastructure and urban commons.

References

Acuto, M. (2018). Global science for city policy. *Science*, *359*(6372), 165–166. https://doi.org/10.1126/science.aao2728

Acuto, M., Parnell, S., & Seto, K. C. (2018). Building a global urban science. *Nature Sustainability*, *1*(1), 2–4. https://doi.org/10.1038/s41893-017-0013-9

Adey, P. (2006). If mobility is everything then it is nothing: Towards a relational politics of (im)mobilities. *Mobilities*, *1*(1), 75–94. https://doi.org/10.1080/17450100500489080

Aguero-Valverde, J., & Jovanis, P. P. (2006). Spatial analysis of fatal and injury crashes in Pennsylvania. *Accident Analysis & Prevention*, *38*(3), 618–625. https://doi.org/10.1016/j.aap.2005.12.006

Ahn, J. W., Yi, M. S., & Shin, D. B. (2013). Study for spatial big data concept and system building. *Spatial Information Research*, *21*(5), 43–51. https://doi.org/10.12672/ksis.2013.21.5.043

Amin, A. (2020). On urban failure. In S. Goldhill (Ed.), *Being urban* (pp. 59–72). Routledge.

Anderson, K., Nafus, D., Rattenbury, T., & Aipperspach, R. (2009). Numbers have qualities too: Experiences with ethno-mining. *Ethnographic Praxis in Industry Conference Proceedings*, *1*, 123–140.

Anderson, T. K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident Analysis & Prevention*, *41*(3), 359–364. https://doi.org/10.1016/j.aap.2008.12.014

Arribas-Bel, D. (2014). Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography*, *49*(May), 45–53. https://doi.org/10.1016/j.apgeog.2013.09.012

Arumugam, S., & Bhargavi, R. (2019). A survey on driving behavior analysis in usage based insurance using big data. *Journal of Big Data*, *6*(1), 1–21. https://doi.org/10.1186/s40537-019-0249-5

Athey, S. (2017). Beyond prediction: Using big data for policy problems. *Science*, *355*(6324), 483–485. https://doi.org/10.1126/science.aal4321

Azmin, M., Jafari, A., Rezaei, N., Bhalla, K., Bose, D., Shahraz, S., Dehghani, M., Niloofar, P., Fatholahi, S., Hedayati, J., Jamshidi, H., & Farzadfar, F. (2018). An approach towards reducing road traffic injuries and improving public health through big data telematics: A randomized controlled trial protocol. *Archives of Iranian Medicine*, *21*(11), 495–501. http://www.aimjournal.ir/Article/aim-3434

Bai, X., Surveyer, A., Elmqvist, T., Gatzweiler, F. W., Güneralp, B., Parnell, S., Prieur-Richard, A.-H., Shrivastava, P., Siri, J. G., Stafford-Smith, M., Toussaint, J.-P., & Webb, R. (2016). Defining and advancing a systems approach for sustainable cities. *Current Opinion in Environmental Sustainability*, *23*(December), 69–78. https://doi.org/10.1016/j.cosust.2016.11.010

Bannister, J., & O'Sullivan, A. (2021). Big data in the city. *Urban Studies*, *58*(15), 3061–3070. https://doi.org/10.1177/00420980211014124

Barnett, C. (2020). The strange case of urban theory. *Cambridge Journal of Regions, Economy and Society*, *13*(3), 443–459. https://doi.org/10.1093/cjres/rsaa026

Barnett, C., & Parnell, S. (2016). Ideas, implementation and indicators: Epistemologies of the post-2015 urban agenda. *Environment and Urbanization*, *28*(1), 87–98. https://doi.org/10.1177/0956247815621473

Batty, M. (2012). Building a science of cities. *Cities*, *29*(Supplement 1), S9–S16. https://doi.org/10.1016/j.cities.2011.11.008

Batty, M. (2013). *The new science of cities*. MIT Press.

Batty, M. (2019). Urban analytics defined. *Environment and Planning B: Urban Analytics and City Science*, *46*(3), 403–405. doi:10.1177/2399808319839494.

Batty, M., Axhausen, K. W., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., Ouzounis, G., & Portugali, Y. (2012). Smart cities of the future. *The European Physical Journal. Special Topics*, *214*(1), 481–518. https://doi.org/10.1140/epjst/e2012-01703-3

Benjamin, R. (2019). Assessing risk, automating racism. *Science*, *366*(6464), 421–422. https://doi.org/10.1126/science.aaz3873

Bernard, H. R. (2002). *Research methods in anthropology: Qualitative and quantitative approaches*. Rowman & Littlefield.

Bibri, S. E., & Krogstie, J. (2017). ICT of the new wave of computing for sustainable urban forms: Their big data and context-aware augmented typologies and design concepts. *Sustainable Cities and Society*, *32*(July), 449–474. https://doi.org/10.1016/j.scs.2017.04.012

Bíl, M., Andrášik, R., & Janoška, Z. (2013). Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation. *Accident Analysis & Prevention*, *55*(June), 265–273. https://doi.org/10.1016/j.aap.2013.03.003

Blok, A., Carlsen, H. B., Jørgensen, T. B., Madsen, M. M., Ralund, S., & Pedersen, M. A. (2017). Stitching together the heterogeneous party: A complementary social data science experiment. *Big Data & Society*, *4*(2), 2053951717736337. https://doi.org/10.1177/2053951717736337

Bornakke, T., & Due, B. L. (2018). Big–thick blending: A method for mixing analytical insights from big and thick data sources. *Big Data & Society*, *5*(1), 2053951718765026. https://doi.org/10.1177/2053951718765026

Boy, J. D., Uitermark, J., & Preis, T. (2016). How to study the city on Instagram. *PloS ONE*, *11*(6), e0158161. https://doi.org/10.1371/journal.pone.0158161

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Cabrera-Arnau, C., Prieto Curiel, R., & Bishop, S. R. (2020). Uncovering the behaviour of road accidents in urban areas. *Royal Society Open Science*, *7*(4), 191739. https://doi.org/10.1098/rsos.191739

Cao, K., Diao, M., & Wu, B. (2019). A big data-based geographically weighted regression model for public housing prices: A case study in Singapore. *Annals of the American Association of Geographers*, *109*(1), 173–186. https://doi.org/10.1080/24694452.2018.1470925

Caprotti, F., & Cowley, R. (2017). Interrogating urban experiments. *Urban Geography*, *38*(9), 1441–1450. https://doi.org/10.1080/02723638.2016.1265870

Chen, F., Chen, S., & Ma, X. (2018). Analysis of hourly crash likelihood using unbalanced panel data mixed logit model and real-time driving environmental big data. *Journal of Safety Research*, *65*(June), 153–159. https://doi.org/10.1016/j.jsr.2018.02.010

Chen, Y., Shu, L., & Wang, L. (2017). Traffic flow prediction with big data: A deep learning based time series model. *2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 1010–1011. Institute of Electrical and Electronics Engineers, Atlanta, GA, USA.

Cinnamon, J., Jones, S. K., & Adger, W. N. (2016). Evidence and future potential of mobile phone data for disease disaster management. *Geoforum*, *75*(October), 253–264. https://doi.org/10.1016/j.geoforum.2016.07.019

Coletta, C., Evans, L., Heaphy, L., & Kitchin, R. (2018). *Creating smart cities*. Routledge.

Cresswell, T. (2006). *On the move: Mobility in the modern Western world*. Taylor & Francis.

Creutzig, F., Lohrey, S., Bai, X., Baklanov, A., Dawson, R., Dhakal, S., Lamb, W. F., McPhearson, T., Minx, J., Munoz, E., & Walsh, B. (2019). Upscaling urban data science for global climate solutions. *Global Sustainability*, *2*(e2). https://doi.org/10.1017/sus.2018.16

Curran, J. (2013). Big data or 'big ethnographic data'? Positioning big data within the ethnographic space. *Ethnographic Praxis in Industry Conference Proceedings*, *2013*, 62–73. Wiley Online Library.

Dameri, R. P., Benevolo, C., Veglianti, E., & Li, Y. (2019). Understanding smart cities as a glocal strategy: A comparison between Italy and China. *Technological Forecasting and Social Change*, *142*(May), 26–41. https://doi.org/10.1016/j.techfore.2018.07.025

Danley, S. (2021). An activist in the field: Social media, ethnography, and community. *Journal of Urban Affairs*, *43*(3), 397–413. https://doi.org/10.1080/07352166.2018.1511797

Datta, A. (2015). New urban utopias of postcolonial India: 'Entrepreneurial urbanization' in Dholera smart city, Gujarat. *Dialogues in Human Geography*, *5*(1), 3–22. https://doi.org/10.1177/2043820614565748

De Oliveira, Á., Campolargo, M., & Martins, M. (2015). Constructing human smart cities. In M. Helfert, K.-H. Krempels, C. Klein, B. Donellan, & O. Guiskhin (Eds.), *Smart cities, green technologies, and intelligent transport systems* (pp. 32–49). Springer.

DeJoy, D. M. (1989). The optimism bias and traffic accident risk perception. *Accident Analysis & Prevention*, *21*(4), 333–340. https://doi.org/10.1016/0001-4575(89)90024-9

Delhomme, P. (1991). Comparing one's driving with others': Assessment of abilities and frequency of offences. Evidence for a superior conformity of self-bias? *Accident Analysis & Prevention*, *23*(6), 493–508. https://doi.org/10.1016/0001-4575(91)90015-W

Di Clemente, R., Luengo-Oroz, M., Travizano, M., Xu, S., Vaitla, B., & González, M. C. (2018). Sequences of purchases in credit card data reveal lifestyles in urban populations. *Nature Communications*, *9*(1), 1–8. https://doi.org/10.1038/s41467-018-05690-8

Doran, D., Severin, K., Gokhale, S., & Dagnino, A. (2016). Social media enabled human sensing for smart cities. *AI Communications*, *29*(1), 57–75. https://doi.org/10.3233/AIC-150683

Dourish, P., & Gómez Cruz, E. (2018). Datafication and data fiction: Narrating data and narrating with data. *Big Data & Society*, *5*(2), 2053951718784083. https://doi.org/10.1177/2053951718784083

Drottenborg, H. (1999). *Aesthetics and safety in traffic environments*. Lund Institute of Technology.

Duminy, J., & Parnell, S. (2020). City science: A chaotic concept–And an enduring imperative. *Planning Theory & Practice*, *21*(4), 648–655. https://doi.org/10.1080/14649357.2020.1802155

Duneier, M., Kasinitz, P., & Murphy, A. (2014). *The urban ethnography reader*. Oxford University Press.

Edwards, J. B. (1996). Weather-related road accidents in England and Wales: A spatial analysis. *Journal of Transport Geography*, *4*(3), 201–212. https://doi.org/10.1016/0966-6923(96)00006-3

Elmqvist, T., Bai, X., Frantzeskaki, N., Griffith, C., Maddox, D., McPhearson, T., Parnell, S., Romero-Lankao, P., Simon, D., & Watkins, M. (editors). (2018). *The urban planet: Knowledge towards sustainable cities*. Cambridge University Press.

Elvik, R., Christensen, P., & Amundsen, A. (2004). *Speed and road accidents: An evaluation of the power model*. Institute of Transport Economics (TOI), Norwegian Centre for Transport Research.

Erdogan, S., Yilmaz, I., Baybura, T., & Gullu, M. (2008). Geographical information systems aided traffic accident analysis system case study: City of Afyonkarahisar. *Accident Analysis & Prevention*, *40*(1), 174–181. https://doi.org/10.1016/j.aap.2007.05.004

Falk, B., & Montgomery, H. (2007). Developing traffic safety interventions from conceptions of risks and accidents. *Transportation Research. Part F, Traffic Psychology and Behaviour*, *10*(5), 414–427. https://doi.org/10.1016/j.trf.2007.04.001

Fan, Z., Liu, C., Cai, D., & Yue, S. (2019). Research on black spot identification of safety in urban traffic accidents based on machine learning method. *Safety Science*, *118*(October), 607–616. https://doi.org/10.1016/j.ssci.2019.05.039

Flyvbjerg, B. (2004). Phronetic planning research: Theoretical and methodological reflections. *Planning Theory & Practice*, *5*(3), 283–306. https://doi.org/10.1080/1464935042000250195

Ford, H. (2014). Big data and small: Collaborations between ethnographers and data scientists. *Big Data & Society*, *1*(2), 205395171454433. https://doi.org/10.1177/2053951714544337

Frantzeskaki, N., McPhearson, T., & Kabisch, N. (2021). Urban sustainability science: Prospects for innovations through a system's perspective, relational and transformations' approaches. *Ambio*, *50*(9), 1650–1658. https://doi.org/10.1007/s13280-021-01521-1

Gabrys, J., Pritchard, H., & Barratt, B. (2016). Just good enough data: Figuring data citizenships through air pollution sensing and data stories. *Big Data & Society*, *3*(2), 205395171667967. https://doi.org/10.1177/2053951716679677

Ge, Q., & Fukuda, D. (2016). Updating origin–destination matrices with aggregated data of GPS traces. *Transportation Research Part C: Emerging Technologies*, *69*(August), 291. https://doi.org/10.1016/j.trc.2016.06.002

Geertz, C. (1973). *The interpretation of cultures*. Basic Books New York.

Girardin, F. (2013). *Insights from network data analysis that yield field observations*. Ethnography Matters. Retrieved November 20, 2019, from https://ethnographymatters.net/blog/2013/04/02/insights-from-networkdata-analysis-that-yield-field-observations/

Goffman, E. (1968). *Asylums: Essays on the social situation of mental patients and other inmates*. AldineTransaction.

Gong, J., Li, S., Ye, X., Peng, Q., & Kudva, S. (2021). Modeling impacts of high-speed rail on urban interaction with social media in China's mainland. *Geo-spatial Information Science*, *24*(4), 638–653. doi:10.1080/10095020.2021.1972771.

Graham, S. (2000). Constructing premium network spaces: Reflections on infrastructure networks and contemporary urban development. *International Journal of Urban and Regional Research*, *24*(1), 183–200. https://doi.org/10.1111/1468-2427.00242

Guan, C., Song, J., Keith, M., Akiyama, Y., Shibasaki, R., & Sato, T. (2020). Delineating urban park catchment areas using mobile phone data: A case study of Tokyo. *Computers, Environment and Urban Systems*, *81*(May), 101474. https://doi.org/10.1016/j.compenvurbsys.2020.101474

Guan, C., Song, J., Keith, M., Zhang, B., Akiyama, Y., Da, L., Shibasaki, R., & Sato, T. (2021). Seasonal variations of park visitor volume and park service area in Tokyo: A mixed-method approach combining big data and field observations. *Urban Forestry and Urban Greening*, *58*(March), 126973. https://doi.org/10.1016/j.ufug.2020.126973

Hammersley, M. (1992). *What's wrong with ethnography? Methodological explorations*. Routledge.

Hammersley, M. (2006). Ethnography: Problems and prospects. *Ethnography and Education*, *1*(1), 3–14. https://doi.org/10.1080/17457820500512697

Hernández-Hernández, A. M., Siqueiros-García, J. M., Robles-Belmont, E., Gershenson, C., & Useche, S. A. (2019). Anger while driving in Mexico City. *PloS ONE*, *14*(9), e0223048. https://doi.org/10.1371/journal.pone.0223048

Hollands, R. G. (2008). Will the real smart city please stand up? Intelligent, progressive or entrepreneurial? *City*, *12*(3), 303–320. https://doi.org/10.1080/13604810802479126

Holmes, B. D., Haglund, K., Beyer, K. M., & Cassidy, L. D. (2019). Qualitative methods of road traffic crash research in low- and middle-income countries: A review. *International Journal of Injury Control and Safety Promotion*, *26*(2), 194–199. https://doi.org/10.1080/17457300.2018.1535512

Horwood, L. J., & Fergusson, D. M. (2000). Drink driving and traffic accidents in young people. *Accident Analysis & Prevention*, *32*(6), 805–814. https://doi.org/10.1016/S0001-4575(00)00005-1

Hu, Q., & Zheng, Y. (2021). Smart city initiatives: A comparative study of American and Chinese cities. *Journal of Urban Affairs*, *43*(4), 504–525. https://doi.org/10.1080/07352166.2019.1694413

Jensen, O. B. (2009). Flows of meaning, cultures of movements–Urban mobility as meaningful everyday life practice. *Mobilities*, *4*(1), 139–158. https://doi.org/10.1080/17450100802658002

Kanarachos, S., Christopoulos, S.-R. G., & Chroneos, A. (2018). Smartphones as an integrated platform for monitoring driver behaviour: The role of sensor fusion and connectivity. *Transportation Research Part C: Emerging Technologies*, *95*(October), 867–882. https://doi.org/10.1016/j.trc.2018.03.023

Kang, W., Oshan, T., Wolf, L. J., Boeing, G., Frias-Martinez, V., Gao, S., Poorthuis, A., & Xu, W. (2019). A roundtable discussion: Defining urban data science. *Environment and Planning B: Urban Analytics and City Science*, *46*(9), 1756–1768. https://doi.org/10.1177/2399808319882826

Karvonen, A., Cvetkovic, V., Herman, P., Johansson, K., Kjellström, H., Molinari, M., & Skoglund, M. (2021). The 'new urban science': Towards the interdisciplinary and transdisciplinary pursuit of sustainable transformations. *Urban Transformations*, *3*(1), 1–13. https://doi.org/10.1186/s42854-021-00028-y

Karvonen, A., & Van Heur, B. (2014). Urban laboratories: Experiments in reworking cities. *International Journal of Urban and Regional Research*, *38*(2), 379–392. https://doi.org/10.1111/1468-2427.12075

Kawlra, G., & Sakamoto, K. (2021). Spatializing urban health vulnerability: An analysis of NYC's critical infrastructure during COVID-19. *Urban Studies*.

Keith, M., O'Clery, N., Parnell, S., & Revi, A. (2020). The future of the future city? The new urban sciences and a PEAK urban interdisciplinary disposition. *Cities*, *105*(October), 102820. https://doi.org/10.1016/j.cities.2020.102820

Kim, H. J., Chae, B. K., & Park, S. B. (2018). Exploring public space through social media: An exploratory case study on the high line New York City. *Urban Design International*, *23*(2), 69–85. https://doi.org/10.1057/s41289-017-0050-z

Kitchin, R. (2014a). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1), 2053951714528481. https://doi.org/10.1177/2053951714528481

Kitchin, R. (2014b). The real-time city? Big data and smart urbanism. *GeoJournal*, *79*(1), 1–14. https://doi.org/10.1007/s10708-013-9516-8

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). *The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data*. United States National Highway Traffic Safety Administration.

Kozinets, R. (2015). *Netnography: Redefined* (2nd ed.). Sage.

Kumar, S., & Toshniwal, D. (2016). A data mining approach to characterize road accident locations. *Journal of Modern Transportation*, *24*(1), 62–72. https://doi.org/10.1007/s40534-016-0095-5

Kwon, O. H., Rhee, W., & Yoon, Y. (2015). Application of classification algorithms for analysis of road safety risk factor dependencies. *Accident Analysis & Prevention*, *75*, 1–15. https://doi.org/10.1016/j.aap.2014.11.005

Laaksonen, S.-M., Nelimarkka, M., Tuokko, M., Marttila, M., Kekkonen, A., & andVilli, M. (2017). Working the fields of big data: Using big-data-augmented online ethnography to study candidate–candidate interaction at election time. *Journal of Information Technology & Politics*, *14*(2), 110–131. https://doi.org/10.1080/19331681.2016.1266981

Lamont, M. (2012). Accidents have no cure! Road death as industrial catastrophe in Eastern Africa. *African Studies*, *71*(2), 174–194. https://doi.org/10.1080/00020184.2012.702964

Leventon, J., Abson, D. J., & Lang, D. J. (2021). Leverage points for sustainability transformations: Nine guiding questions for sustainability science and practice. *Sustainability Science*, *16*(3), 721–726. https://doi.org/10.1007/s11625-021-00961-8

Lewis, C., & Symons, J. (2017). *Realising the city: Urban ethnography in Manchester*. Manchester University Press.

Li, S., & Yang, B. (2021). How important are the park size and shape to a park system's performance? An exploration with big data in Tucson, Arizona, USA. *Socio-Ecological Practice Research*, *3*(3), 281–291. https://doi.org/10.1007/s42532-021-00086-3

Li, W., Batty, M., & Goodchild, M. F. (2020). Real-time GIS for smart cities. *International Journal of Geographical Information Science*, *34*(2), 311–324. https://doi.org/10.1080/13658816.2019.1673397

Liu, C., Liu, Z., & Guan, C. (2021). The impacts of the built environment on the incidence rate of COVID-19: A case study of King County, Washington. *Sustainable Cities & Society*, *74*(November), 103144. https://doi.org/10.1016/j.scs.2021.103144

Lyon, C., Brown, S., Vanlaar, W., & Robertson, R. (2021). Prevalence and trends of distracted driving in Canada. *Journal of Safety Research*, *76*(February), 118–126. https://doi.org/10.1016/j.jsr.2020.12.005

Magro, J. L. (2018). Resistance identities and language choice in Instagram among Hispanic urban artists in Da DMV: Big data and a mixed-method. *Education for Information*, *34*(3), 215–238. https://doi.org/10.3233/EFI-180199

Marsden, G., & Reardon, L. (2017). Questions of governance: Rethinking the study of transportation policy. *Transportation Research Part A: Policy and Practice*, *101*(July), 238–251. https://doi.org/10.1016/j.tra.2017.05.008

McPhearson, T., Pickett, S. T., Grimm, N. B., Niemelä, J., Alberti, M., Elmqvist, T., Weber, C., Haase, D., Breuste, J., & Qureshi, S. (2016). Advancing urban ecology toward a science of cities. *BioScience*, *66*(3), 198–212. https://doi.org/10.1093/biosci/biw002

Mohan, D., Jha, A., & Chauhan, S. S. (2021). Future of road safety and SDG 3.6 goals in six Indian cities. *IATSS Research*, *45*(1), 12–18. https://doi.org/10.1016/j.iatssr.2021.01.004

Mora, L., Deakin, M., Zhang, X., Batty, M., de Jong, M., Santi, P., & Appio, F. P. (2021). Assembling sustainable smart city transitions: An interdisciplinary theoretical perspective. *Journal of Urban Technology*, *28*(1–2), 1–27. https://doi.org/10.1080/10630732.2020.1834831

Moran, M., Baron-Epel, O., & Assi, N. (2010). Causes of road accidents as perceived by Arabs in Israel: A qualitative study. *Transportation Research. Part F, Traffic Psychology and Behaviour*, *13*(6), 377–387. https://doi.org/10.1016/j.trf.2010.07.001

Najaf, P., Thill, J.-C., Zhang, W., & Fields, M. G. (2018). City-level urban form and traffic safety: A structural equation modeling analysis of direct and indirect effects. *Journal of Transport Geography*, *69*(May), 257–270. https://doi.org/10.1016/j.jtrangeo.2018.05.003

Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., Mascolo, C., & Añel, J. A. (2012). A tale of many cities: Universal patterns in human urban mobility. *PloS ONE*, *7*(5), e37027. https://doi.org/10.1371/journal.pone.0037027

O'Halloran, K. L., Tan, S., Pham, D.-S., Bateman, J., & Vande Moere, A. (2018). A digital mixed methods research design: Integrating multimodal analysis with data mining and information visualization for big data analytics. *Journal of Mixed Methods Research*, *12*(1), 11–30. https://doi.org/10.1177/1558689816651015

O'Halloran, K. L., Tan, S., Wignell, P., Bateman, J. A., Pham, D.-S., Grossman, M., & Moere, A. V. (2019). Interpreting text and image relations in violent extremist discourse: A mixed methods approach for big data analytics. *Terrorism and Political Violence*, *31*(3), 454–474. https://doi.org/10.1080/09546553.2016.1233871

Ortman, S. G., Lobo, J., Smith, M. E., & Biehl, P. F. (2020). Cities: Complexity, theory and history. *PloS ONE*, *15*(12), e0243621. https://doi.org/10.1371/journal.pone.0243621

Pappalardo, L., Pedreschi, D., Smoreda, Z., & Giannotti, F. (2015a). Using big data to study the link between human mobility and socio-economic development. *2015 IEEE International Conference on Big Data (Big Data)*, Santa Clara, CA, IEEE, 871–878.

Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., & Barabási, A.-L. (2015b). Returners and explorers dichotomy in human mobility. *Nature Communications*, *6*(1), 8166. https://doi.org/10.1038/ncomms9166

Pardo, I., & Prato, G. B. (2018). Introduction: Urban ethnography matters— Analytical strength, theoretical value and significance to society. In I. Pardo & G. B. Prato (Eds.), *The Palgrave handbook of urban ethnography* (pp. 1–19). Springer.

Parnell, S., & Robinson, J. (2017). The global urban: Difference and complexity in urban studies and the science of cities. In S. Hall & R. Burdett (Eds.), *The SAGE handbook of the 21st century city* (pp. 13–31). SAGE.

Pilkington, P., Bird, E., Gray, S., Towner, E., Weld, S., & McKibben, M.-A. (2014). Understanding the social context of fatal road traffic collisions among young people: A qualitative analysis of narrative text in coroners' records. *BMC Public Health*, *14*(1), 78. https://doi.org/10.1186/1471-2458-14-78

Prasannakumar, V., Vijith, H., Charutha, R., & Geetha, N. (2011). Spatiotemporal clustering of road accidents: GIS based analysis and assessment. *Procedia - Social and Behavioral Sciences*, *21*, 317–325. https://doi.org/10.1016/j.sbspro.2011.07.020

Pretnar, A., & Podjed, D. (2019). Data mining workspace sensors: A new approach to anthropology. *Prispevki Za Novejšo Zgodovino*, *59*(1), 179–197. https://www.ceeol.com/search/article-detail?id=886566

Prieto Curiel, R., González Ramírez, H., Bishop, S. R., & Deng, Y. (2018). A novel rare event approach to measure the randomness and concentration of road accidents. *PloS ONE*, *13*(8), e0201890. https://doi.org/10.1371/journal.pone.0201890

Prieto Curiel, R., Patino, J. E., Duque, J., O'Clery, N., & Yuan, Q. (2021). The heartbeat of the city. *PLoS ONE*, *16*(2), e0246714. https://doi.org/10.1371/journal.pone.0246714

Rabari, C., & Storper, M. (2015). The digital skin of cities: Urban theory and research in the age of the sensored and metered city, ubiquitous computing and big data. *Cambridge Journal of Regions, Economy and Society*, *8*(1), 27–42. https://doi.org/10.1093/cjres/rsu021

Ramos, P., Diez, E., Pérez, K., Rodriguez-Martos, A., Brugal, M. T., & andVillalbí, J. R. (2008). Young people's perceptions of traffic injury risks, prevention and enforcement measures: A qualitative study. *Accident Analysis & Prevention*, *40*(4), 1313–1319. https://doi.org/10.1016/j.aap.2008.02.001

Reese, L. (2014). The present and future of urban affairs research. *Journal of Urban Affairs*, *36*(sup2), 543–550. https://doi.org/10.1111/juaf.12143

Rinzivillo, S., Gabrielli, L., Nanni, M., Pappalardo, L., Pedreschi, D., & Giannotti, F. (2014). The purpose of motion: Learning activities from individual mobility networks. *2014 International Conference on Data Science and Advanced Analytics (DSAA)*, Shanghai, China, IEEE, 312–318.

Rubrichi, S., Smoreda, Z., & Musolesi, M. (2018). A comparison of spatial-based targeted disease mitigation strategies using mobile phone data. *EPJ Data Science*, *7*(1), 17. https://doi.org/10.1140/epjds/s13688-018-0145-9

Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. *2013 International Conference on Collaboration Technologies and Systems (CTS)*, San Diego, CA, 42–47. IEEE.

Salah, A. A., Pentland, A., Lepri, B., Letouze, E., Vinck, P., de Montjoye, Y.-A., Dong, X., & Dagdelen, O. (2018). Data for refugees: The D4R challenge on mobility of Syrian refugees in Turkey. *arXiv*. https://arxiv.org/abs/1807.00523

Sanusi, A. A., & Emmelin, M. (2015). Commercial motorcycle drivers' perceptions of risk and road safety in urban Nigeria: An explorative study. *International Journal of Injury Control and Safety Promotion*, *22*(4), 328–339. https://doi.org/10.1080/17457300.2014.909499

Savolainen, P. T., Mannering, F. L., Lord, D., & Quddus, M. A. (2011). The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. *Accident Analysis & Prevention*, *43*(5), 1666–1676. https://doi.org/10.1016/j.aap.2011.03.025

Schensul, J. J., and LeCompte, M. D. (Ed.). (2012). *Essential ethnographic methods: A mixed methods approach* (Vol. 3). AltaMira Press.

Schneider, C. M., Belik, V., Couronné, T., Smoreda, Z., & González, M. C. (2013). Unravelling daily human mobility motifs. *Journal of the Royal Society Interface*, *10*(84), 20130246. https://doi.org/10.1098/rsif.2013.0246

Schuurman, N., Cinnamon, J., Crooks, V. A., & Hameed, S. M. (2009). Pedestrian injury and the built environment: An environmental scan of hotspots. *BMC Public Health*, *9*(1), 233. https://doi.org/10.1186/1471-2458-9-233

Schwanen, T. (2017). Geographies of transport II: Reconciling the general and the particular. *Progress in Human Geography*, *41*(3), 355–364. https://doi.org/10.1177/0309132516628259

Shaw, S.-L., Tsou, M.-H., & Ye, X. (2016). Human dynamics in the mobile and big data era. *International Journal of Geographical Information Science*, *30*(9), 1687–1693. https://doi.org/10.1080/13658816.2016.1164317

Sheller, M., & Urry, J. (2006). The new mobilities paradigm. *Environment and Planning A: Economy and Space*, *38*(2), 207–226. https://doi.org/10.1068/a37268

Smith, M. E., Lobo, J., Peeples, M. A., York, A. M., Stanley, B. W., Crawford, K. A., Gauthier, N., & Huster, A. C. (2021). The persistence of ancient settlements and urban sustainability. *Proceedings of the National Academy of Sciences*, *118*(20), e2018155118. https://doi.org/10.1073/pnas.2018155118

Smith, R. M., Pathak, P. A., & Agrawal, G. (2019). India's "smart" cities mission: A preliminary examination into India's newest urban development policy. *Journal of Urban Affairs*, *41*(4), 518–534. https://doi.org/10.1080/07352166.2018.1468221

Steenberghen, T., Aerts, K., & Thomas, I. (2010). Spatial clustering of events on a network. *Journal of Transport Geography*, *18*(3), 411–418. https://doi.org/10.1016/j.jtrangeo.2009.08.005

Stehle, S., & Kitchin, R. (2020). Real-time and archival data visualization techniques in city dashboards. *International Journal of Geographical Information Science*, *34*(2), 344–366. https://doi.org/10.1080/13658816.2019.1594823

Stewart, M. A. (1995). Effective physician-patient communication and health outcomes: A review. *CMAJ: Canadian Medical Association Journal*, *152*(9), 1423. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1337906/

Sun, Y., Shao, Y., & Chan, E. H. (2020). Co-visitation network in tourism-driven peri-urban area based on social media analytics: A case study in Shenzhen, China. *Landscape and Urban Planning*, *204*(December), 103934. https://doi.org/10.1016/j.landurbplan.2020.103934

Tan, M., & Guan, C. (2021). Are people happier in locations of high property value? Spatial temporal analytics of activity frequency, public sentiment and housing price using Twitter data. *Applied Geography*, *132*(July), 102474. https://doi.org/10.1016/j.apgeog.2021.102474

Taylor, L. (2021). The taming of chaos: Optimal cities and the state of the art in urban systems research. *Urban Studies*, *58*(15), 3196–3202. https://doi.org/10.1177/00420980211012838

Taylor, L., & Richter, C. (2017). The power of smart solutions: Knowledge, citizenship, and the datafication of Bangalore's water supply. *Television & New Media*, *18*(8), 721–733. https://doi.org/10.1177/1527476417690028

Thomas, I. (1996). Spatial data aggregation: Exploratory analysis of road accidents. *Accident Analysis & Prevention*, *28*(2), 251–264. https://doi.org/10.1016/0001-4575(95)00067-4

Tricarico, L., Jones, Z. M., & Daldanise, G. (2020). Platform spaces: When culture and the arts intersect territorial development and social innovation, a view from the Italian context. *Journal of Urban Affairs*. https://doi.org/10.1080/07352166.2020.1808007

Urry, J. (2016). *Mobilities: New perspectives on transport and society*. Routledge.

Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society*, *12*(2), 197–208. https://doi.org/10.24908/ss.v12i2.4776

Vanolo, A. (2014). Smartmentality: The smart city as disciplinary strategy. *Urban Studies*, *51*(5), 883–898. https://doi.org/10.1177/0042098013494427

Verloo, N., & Bertolini, L. (2020). *Seeing the city: Interdisciplinary perspectives on the study of the urban*. Amsterdam University Press.

Wang, M., & Vermeulen, F. (2021). Life between buildings from a street view image: What do big data analytics reveal about neighborhood organizational vitality? *Urban Studies*, *58*(15), 3118–3139. https://doi.org/10.1177/0042098020957198

Wang, T. (2013). *Big data needs thick data*. Ethnography Matters. Retrieved October 1, 2019, from http://ethnographymatters.net/blog/2013/05/13/big-data-needs-thickdata/

Watters, S. E., & Beck, K. H. (2016). A qualitative study of college students' perceptions of risky driving and social influences. *Traffic Injury Prevention*, *17*(2), 122–127. https://doi.org/10.1080/15389588.2015.1045063

Widhalm, P., Yang, Y., Ulm, M., Athavale, S., & González, M. C. (2015). Discovering urban activity patterns in cell phone data. *Transportation*, *42*(4), 597–623. https://doi.org/10.1007/s11116-015-9598-x

Williams, J. C., Anderson, N., Mathis, M., Sanford, III, E., Eugene, J., & Isom, J. (2020). Colorblind algorithms: Racism in the era of COVID-19. *Journal of the National Medical Association*, *112*(5), 550–552. https://doi.org/10.1016/j.jnma.2020.05.010

Williamson, A., Lombardi, D. A., Folkard, S., Stutts, J., Courtney, T. K., & Connor, J. L. (2011). The link between fatigue and safety. *Accident Analysis & Prevention*, *43*(2), 498–515. https://doi.org/10.1016/j.aap.2009.11.011

Wilson, R., Zu Erbach-Schoenberg, E., Albert, M., Power, D., Tudge, S., Gonzalez, M., Guthrie, S., Chamberlain, H., Brooks, C., Hughes, C., Pitonakova, L., Buckee, C., Lu, X., Wetter, E., Tatem, A., & Bengtsson, L. (2016). Rapid and near real-time assessments of population displacement using mobile phone data following disasters: The 2015 Nepal earthquake. *PLoS Cur- Rents*, *8*(February). doi:10.1371/currents.dis.d073fbece328e4c39087bc086d694b5c.

World Health Organization. (2018). *Global status report on road safety 2018*. World Health Organization, Geneva.

Xie, K., Ozbay, K., Kurkcu, A., & Yang, H. (2017). Analysis of traffic crashes involving pedestrians using big data: Investigation of contributing factors and identification of hotspots. *Risk Analysis*, *37*(8), 1459–1476. https://doi.org/10.1111/risa.12785

Yao, Z., Yang, J., Liu, J., Keith, M., & Guan, C. (2021). Comparing tweet sentiments in megacities using machine learning techniques: In the midst of COVID-19. *Cities*, *116*(September), 103273. https://doi.org/10.1016/j.cities.2021.103273

Zhou, X., Chen, Z., Yeh, A. G., & Yue, Y. (2021). Workplace segregation of rural migrants in urban China: A case study of Shenzhen using cellphone big data. *Environment and Planning B: Urban Analytics and City Science*, *48*(1), 25–42. https://doi.org/10.1177/2399808319846903

Zufiria, P. J., Pastor-Escuredo, D., Úbeda-Medina, L., Hernandez-Medina, M. A., Barriales-Valbuena, I., Morales, A. J., Jacques, D. C., Nkwambi, W., Diop, M. B., Quinn, J., Hidalgo-Sanchís, P., & Luengo-Oroz, M. (2018). Identifying seasonal mobility profiles from anonymized and aggregated mobile phone data. Application in food security. *PloS ONE*, *13*(4), e0195714. https://doi.org/10.1371/journal.pone.0195714