Urban Data Science

TINA LAW and JOSCHA LEGEWIE

Abstract

Data on urban life are more accessible today than ever before. New sources of "big data" such as 311 requests, recorded police activity, digitized student records, and social media capture urban life on an unprecedented temporal and geographical scale. Combined with new and improved computational social science methods for harnessing data, they promise to change urban research in important ways. In this essay, we outline urban data science—an emerging, interdisciplinary approach to studying urban life using big data and computational social science methods. We discuss three key innovations that this approach offers for urban research: (i) a broader and more multifaceted definition of neighborhood activity, (ii) greater knowledge on the role of socio-spatial interdependencies in urban life, and (iii) more dynamic understandings of urban issues and policies. We conclude by highlighting some challenges that urban scholars must collaboratively address as they engage in this new urban data science.

INTRODUCTION

In the late nineteenth century, Du Bois (1996) spent over a year painstakingly conducting door-to-door surveys, mapping physical and social conditions, and assembling archival and census data in order to illustrate—with unprecedented empirical detail—daily life in the historically black Seventh Ward and the city of Philadelphia more broadly. Fast forward to the early twenty-first century and data on cities and neighborhoods are more accessible than ever. Large-scale, digitized data—or "big data"—on urban life abound (Lazer *et al.*, 2009). Digitized administrative data sources regularly capture important aspects of daily life in cities and neighborhoods, such as residents' requests for city services, instances of crime, and students' progress in schools. Geo-tagged social media data track how individuals regularly make use of and move about their neighborhoods and cities. Google Street View and other digital data sources document physical conditions in neighborhoods and cities on an ongoing basis. These new sources of data—as well as the new and improved computational methods that enable

Emerging Trends in the Social and Behavioral Sciences.
Robert A. Scott and Marlis Buchmann (General Editors) with Stephen Kosslyn (Consulting Editor).
 ISBN 978-1-118-90077-2.

these data to be effectively used—offer important opportunities for urban research. But what exactly does big data and computational social science mean for the study of cities and urban life? In this essay, we outline *urban data science*—an emerging, interdisciplinary approach that engages big data and computational social science methods to study urban life. This approach offers three key innovations for urban research: (i) a broader and more multifaceted definition of neighborhood activity, (ii) greater knowledge on the role of socio-spatial interdependencies in urban life, and (iii) more dynamic understandings of urban issues and policies. We discuss each of these key innovations of urban data science and provide examples from recent research in sociology, criminology, political science, urban planning, geography, and communications. We conclude by highlighting some of the challenges that urban scholars must collaboratively address as they engage in this new urban data science.

NEIGHBORHOOD ACTIVITY AS BROAD AND MULTIFACETED

As Jane Jacobs (1961) underscored, the everyday activities of neighbors are the driving force of urban life. As such, an important task for urban scholars is to accurately and meaningfully understand what it is that neighbors do (or do not do). Urban scholars employ diverse data and methods toward this end. Urban ethnographers provide rich accounts of neighborhood life based on their on-the-ground observations and interactions, while quantitative researchers largely rely on census and survey data to illustrate how life differs across neighborhoods.

Big data and computational social science, however, present a sea change in the ability of urban scholars to understand how daily life unfolds in neighborhoods across the world. Indeed, there has never been more data on everyday neighborhood activity than now. Administrative data sources such as municipal 311 or constituent relationship management (CRM) systems provide a detailed (but imperfect) look at citizens' needs and the services available to them (O'Brien, Sampson, & Winship, 2015; Minkoff, 2016). New technologies involving smartphone-based global positioning system (GPS) tracking make it possible to collect data on individuals' routine activities within and beyond their neighborhoods (Browning, Calder, Ford, Boettner, Smith, & Haynie, 2017). Social media data are also often geo-tagged, allowing for the study of online interactions occurring within and across neighborhoods (Golder & Macy, 2014).

These new, "readymade" (Salganik, 2017, pp. 6–8) data sources capture neighborhood activities on an unprecedented temporal and geographical scale. Combined with new and improved computational social science methods for harnessing data, they have the potential to redefine neighborhood activity in a broader and more multifaceted way—specifically by allowing for novel measurements of neighbor behaviors and interactions. Several studies illustrate this potential. For example, Legewie and Schaeffer (2016) use 311 data from New York City to study when and where citizens complain about their neighbors making noise, drinking in public, or blocking their driveway. The data make it possible to identify and analyze these "more subtle forms of conflict that are a defining aspect of everyday life" in urban neighborhoods (Legewie & Schaeffer, 2016, p. 138). Along similar lines, O'Brien, Sampson, and Winship (2015, p. 111 italics in original) use Boston CRM requests to measure the important concept of neighborhood physical disorder in a new way that captures the "two distinct but related aspects" of "*private neglect*" and "*public denigration*."

Moreover, the advent of big data and computational social science methods may alter the traditional definition of neighborhood activity by allowing for measurement of digital neighborhood activity and other novel forms of neighbor behaviors and interactions. As more and more aspects of social life take place online or are mediated by web-based technologies (Golder & Macy, 2014), neighborhood social life also increasingly involves digital activities. For example, Goodspeed (2017, pp. 12–13) notes that contemporary neighborhood life regularly involves many forms of online or Internet-mediated activities, such as coordinating activities with neighbors via NextDoor or other social media platforms, submitting requests for city services via smartphone apps, or finding places to eat via Yelp.com or other websites. Similarly, Lane (2016) points out that urban street life now consists of not just physical interactions between individuals on sidewalks but also online interactions between individuals via social media. Although research on digital neighborhood activity is still nascent, it is clear that big data and computational social science methods will be instrumental in learning about these new types of neighborhood activity, as well as for furthering understanding of long-observed types of neighborhood activity.

URBAN LIFE AS INTERDEPENDENT

Urban scholars have long recognized that cities are fundamentally governed by interdependent social and spatial processes. From this perspective, it is important to understand not only the individuals and groups that live in urban spaces but also how they relate to and affect one another. Likewise, it is important to understand how physical spaces within cities are linked. However, empirically studying these interdependencies in urban life is challenging due to data and methodological constraints. As a result, urban research tends to focus on single neighborhoods instead of "higher order" spatial structures and extralocal effects, as well as single social groups or networks instead of multiple, interacting social groups or networks (Sampson, 2012, pp. 238, 329–30). Even less common are urban studies that analyze social and spatial processes *simultaneously* (Adams, Faust, & Lovasi, 2012; Papachristos, Hureau, & Braga, 2013). In particular, most quantitative urban research treats neighborhoods as independent, isolated "islands" without considering the broader socio-spatial structures and processes in which they are inextricably embedded.

Recent advances in big data and computational social science methods provide new opportunities to directly observe and study social and spatial interdependencies in urban life—namely through social network analysis. Although network science has existed for many years, the modeling and analysis of diverse types of social networks is more feasible today than ever before. Data on social interactions are more readily available, and the computationally intensive nature of social network analysis is aided by increasingly powerful computers, parallel computing, and cloud-based data storage (Golder & Macy, 2014). The study of urban social networks in particular is thriving due to continued progress in the development of theoretical and statistical tools that enable simultaneous analysis of social and spatial contexts (Adams *et al.*, 2012), as well as greater availability of geocoded data on urban social activities (Minkoff, 2016; O'Brien *et al.*, 2015).

The flourishing study of urban social networks may help to illuminate the social and spatial interdependencies that organize urban life by providing more in-depth knowledge on how neighbors' use of shared space mediates their social ties. Two recent studies illustrate this potential. Browning, Calder, Soller, Jackson, and Dirlam (2017) leverage social network analysis and spatial analysis to study how the routine activities of neighbors contributes to neighborhood social organization. Using data from the Los Angeles Family and Neighborhood Survey, they construct several neighborhood-level "ecological networks" based on spatial overlap in the geocoded routine activities of neighbors. By analyzing these networks, they find that residents who live in neighborhoods that are better internally connected (in terms of both quantity and strength of relationships) experience higher levels of neighborhood social organization—a key mediator of important life outcomes. Another example is Hipp, Butts, Acton, Nagle, and Boessen's (2013) study of the well-documented relationship between neighborhood social networks and crime. Using simulated network data, they find that the relationship between social networks and crime is indeed important but rather complex within the context of urban neighborhoods: crime rates tend to be higher in neighborhoods where neighbors are better connected *to each other* (as measured by block-level tie density or the extent to which all neighbors on a block are connected to each other) but lower in neighborhoods where neighbors are better connected *as individuals* with social ties within and beyond their own neighborhoods (as measured by neighbor mean degree or neighbors' average number of social ties). By examining how neighbors share space, both studies are able to understand not only how neighbors are socially connected but also how these social connections can, in turn, affect important life outcomes.

Research on urban social networks using big data and computational social science methods may also offer new insights into enduring urban issues—particularly on how these issues are socially and spatially transmitted. In a recent study, Papachristos *et al.* (2013) integrate methods from social network analysis and spatial analysis to examine gang violence in Boston and Chicago. They use police records to construct citywide networks of gang violence where nodes represent gangs and ties represent exchanges of gun violence. Among other findings, the study highlights that gang violence is more likely to occur between gangs with adjacent "turf," which underscores the need to identify not just where gangs are located but also where they are located in relation to their rivals in order to understand the socio-spatial flow of gang violence. In related research, Legewie and Schaeffer (2016) and Legewie (2018) introduce methods to measure neighborhood boundaries defined as abrupt transitions in the socio-demographic composition of neighborhoods. They show that this relational aspect of the socio-spatial structure is related to neighborhood conflict and crime. In another study, Bastomski, Brazil, and Papachristos (2017) examine neighborhood co-offending networks and violent crime in Chicago. They use arrest records to connect neighborhoods in the city through co-offending ties, and then use k-core decomposition techniques to measure the extent to which neighborhoods are embedded in this citywide network. The study finds that more structurally embedded neighborhoods experience higher rates of violent crime, meaning that violent crime is contingent not just on a neighborhood's internal characteristics as is commonly understood but also on the extent to which it is enmeshed in the citywide co-offending network. These studies attest to the value of using social network analysis to elucidate the interdependent social and spatial processes that underlie complex issues of urban inequality. More broadly, the research highlighted in this section shows that the promising study of urban social networks will continue to push the bounds of urban research as it evolves and finds new ways to leverage big data and computational social science methods to understand urban life more fully.

CITIES AND NEIGHBORHOODS AS DYNAMIC ENTITIES

For urban scholars, an enduring challenge has been the study of cities and neighborhoods over time. While there has been strong and long-standing interest to move beyond "static," cross-sectional studies of cities and neighborhoods, there is little to no longitudinal data on urban processes and few time- and cost-effective ways to collect this type of data (Kirk & Laub, 2010; Sampson, 2012). Indeed, Kirk and Laub (2010, p. 444) describe the problem plainly: "virtually no quantitative data exist that measure changes in neighborhood social and cultural processes over time (e.g., informal social control, fear of crime and disorder, perceptions of the law)."

The advent of big data and computational social science methods, however, represents a potentially pivotal step forward in the study of cities and neighborhoods as it relates to temporal context. The continuous and spatial nature of many administrative and corporate data systems makes it much easier to ascertain longitudinal data on cities and neighborhoods. These data include municipal 311 or CRM requests (Minkoff, 2016; O'Brien *et al.*, 2015), 911 calls (Desmond, Papachristos, & Kirk, 2016), recorded police activity (Legewie, 2016), digitized student records (Legewie & Fagan, 2018), and even Google Street View images (Hwang & Sampson, 2014). Historical data on key urban events and actors are also more readily available, especially in formats conducive to data analysis (Smith & Papachristos, 2016). In addition, new and improved computational social science methods enable these longitudinal data to be used in innovative ways.

This new bevy of data and methodological resources provides many important opportunities for urban research, particularly in terms of studying how discrete actions affect cities and neighborhoods over time. Several recent studies demonstrate the promise of leveraging big data and computational social science methods in this way. For example, Legewie (2016) uses data on over three million "Stop, Question, and Frisk" operations in New York City to examine how incidents of violence against police officers may trigger racial bias in the use of police force. Using continuously recorded police activity data and a quasi-experimental design, Legewie (2016) finds that there is a marked increase in use of force against black residents following shootings of police officers by black suspects. However, the use of force against white and Hispanic residents remains the same, and there is no comparable effect for similar cases involving Hispanic and white suspects. Another study by Desmond *et al.* (2016) uses a similar approach to examine how incidents of police violence against unarmed black men may contribute to legal cynicism. Making use of the 911 system in Milwaukee, they find that 911 calls decrease substantially in black neighborhoods following incidents of police-perpetrated violence.

In addition to examining the role of discrete and often unexpected events, urban big data systems allow for retrospective evaluations of urban policy. Legewie and Fagan (2018), for example, use administrative data on millions of students from New York City public schools and detailed information on crime, arrests, and police stops from the New York Police Department (NYPD) to examine the effect of Operation Impact on the educational performance of minority youth. Under Operation Impact, the NYPD saturated high crime areas with additional police officers with the mission to engage in aggressive order-maintenance policing. The findings show that exposure to police surges can harm African-American boys' educational performance and therefore contribute to the racial achievement gap. More broadly, the study demonstrates that "always-on" big data systems can help scholars to "*travel back in time*" (Salganik, 2017, p. 22 italics in original) and evaluate the consequences of urban policy effectively and at low cost.

Moreover, new longitudinal data and computational social science methods allow urban scholars to study how long-term processes unfold in and affect cities and neighborhoods. Along these lines, a recent study by Hwang and Sampson (2014) uses images from Google Street View to explore how the process of gentrification unfolds over time in different Chicago neighborhoods. Using this new data source, they find that race mediates the process of gentrification: neighborhoods with large numbers of black and Hispanic residents are less likely to gentrify even if they possess other characteristics typically conducive to gentrification, such as geographical proximity to already gentrified neighborhoods. In another recent study, Delmelle (2016, p. 36) applies new and improved computational methods—specifically, clustering procedures and a sequential pattern mining algorithm based on optimal matching distance—to traditional census data in order to develop a "typology of neighborhood trajectories" for Los Angeles and Chicago. Delmelle (2016, p. 41) classifies all Los Angeles and Chicago neighborhoods into one of several neighborhood socioeconomic types for five points in time between 1970 and 2010 (e.g., "Newer suburban," "Older, stable suburban," "Blue collar," "Struggling," and "Young urban" types for Chicago), and then clusters the five-component sequences into common types of neighborhood socioeconomic trajectories. By examining the socioeconomic development of neighborhoods in this way, the study shows that urban neighborhoods follow many different trajectories—both within and across cities. Taken together, these studies demonstrate that big data and computational social science methods now make it possible to explore how cities and neighborhoods evolve over time and they showcase innovative ways to move toward more dynamic analyses of urban life.

CONCLUSION

Big data and computational social science methods have the potential to significantly transform and strengthen research on cities and urban life—whether it is by redefining neighborhood activity in a broader and more multifaceted way, illuminating important social and spatial interdependencies that organize urban life, or enabling more dynamic understandings of key urban issues and policies. Advancements in these core dimensions of urban research will allow scholars to empirically test long-standing theories about urban life, pose new questions, and better inform urban policymaking. As the studies discussed in this essay show, big data and computational social science methods have already generated new insights into some of the most important theoretical issues in urban research, such as how neighborhood social (dis)organization originates, what shapes citizens' trust in local government, and where and when gentrification happens.

At the same time, the advent of big data and computational social science methods introduces many new challenges that urban scholars must collaboratively address in order to effectively use these new tools and resources. Indeed, it is essential to consider both the advantages and limitations of these ongoing changes to urban research. In particular, we anticipate three main challenges for urban data science: (i) engaging in ongoing discussions about data access, (ii) understanding how big data are generated, and (iii) finding innovative ways to validate novel data.

First, data accessibility enables urban research and social science research more broadly to be transparent, open, and reproducible. These principles are key for scientific work and build confidence in research findings (Nosek *et al.*, 2015). However, legal, business, and ethical considerations prevent open access to many "big data" sources (Salganik, 2017, pp. 27–29; see also Connelly, Playford, Gayle, & Dibben, 2016). These restrictions are required by federal, state, and local laws, data use agreements, and IRB protocols, and they are important for protecting privacy and business interests. Urban scholars in particular should expect to encounter issues related to data accessibility as urban data become more granular and as companies increasingly collect and privatize data on urban life. Moving forward, urban scholars and other social science researchers who use big data and computational social science methods need to develop shared standards that protect privacy and business interests while maintaining transparency, openness, and reproducibility. To do so, they must engage policymakers, business leaders, and the public in an ongoing dialogue on how privacy and knowledge can both be prioritized in this new research landscape.

Second, in order to use big data to produce new knowledge about urban issues, researchers need to know how exactly these data are generated. Social science research traditionally relies on tailored and well-documented data-generating processes, using measurement tools that are designed to capture theoretical concepts of interest and that build on representative samples or other clear sampling frames based on well-defined populations.

However, the "readymade" nature of many big data sources makes it so that researchers are often not involved in the data collection process—meaning that they frequently have no influence on measurement and sampling and potentially limited information on how data were collected and who comprises the study population (Connelly *et al.*, 2016; Salganik, 2017). These basic gaps in knowledge make it difficult for researchers to assess the efficacy of their research designs and to identify and address limitations in their work (Connelly *et al.,* 2016; Salganik, 2017). For example, uncertainty about whether a sample is representative can make it challenging to answer most descriptive research questions, and even research based on within-sample comparisons that reveals causal mechanisms or other relations and processes faces challenges for the external validity of their findings. Having limited information on how data are generated is especially problematic for urban scholars as their research often requires data that are precisely attributed to specific units of analysis such as neighborhoods or blocks. As urban scholars and other social science researchers increasingly adopt big data, it will be important to invest resources in studying the data themselves (Connelly *et al.,* 2016; Salganik, 2017).

Third, new sources of urban big data often face an inherent tension: while many of these data sources provide new information on previously under-studied or undocumented social activities, interactions, and processes, these data are—by definition—new and therefore may be difficult to validate with external data sources (O'Brien *et al.,* 2015). In fact, some of the most interesting data in 311 or CRM databases and other big data sources are also some of the "most difficult to validate" (O'Brien *et al.,* 2015, p. 138). Validating urban big data is especially challenging given that there may be few sources of data for a specific city or neighborhood, and cities and neighborhoods are often idiosyncratic in their histories and socio-spatial characteristics. As such, urban scholars and other social science researchers who use big data will need to find innovative ways to validate novel data sources. Addressing this challenge will encourage greater use of big data in urban research and other social science research, as well as ensure engagement between urban data science and long-standing theoretical and methodological traditions in urban research.

These challenges will become increasingly important as more and more urban scholars embrace big data and computational social science methods. In highlighting key limitations, they also powerfully illustrate an important conclusion: big data and computational social science methods are here not to replace but instead to supplement existing data and methods. Indeed, Glaeser *et al.* (2018, p. 114) note that "big data will not solve large urban social science problems on its own." Instead, these new data sources shine in combination with traditional forms of data collection (Glaeser *et al.,* 2018).

Small (2017, p. 176) similarly points out that while computational social science methods such as social network analysis are adept at answering certain key questions (e.g., what is the structure of a social network), they are ill-equipped to address others (e.g., when, how, and why do people activate—or do not activate—their social networks). Therefore, it is evident that the advent of big data and computational social science methods does not render obsolete long-standing data sources and methodological traditions in urban research. If anything, they underscore the importance of integrating different data and methods, and they offer new opportunities for collaboration within and across disciplines.

REFERENCES

Adams, J., Faust, K., & Lovasi, G. S. (2012). Capturing context: Integrating spatial and social network analyses. *Social Networks*, *34*(1), 1–5.

Bastomski, S., Brazil, N., & Papachristos, A. V. (2017). Neighborhood co-offending networks, structural embeddedness, and violent crime in Chicago. *Social Networks*, *51*(October), 23–39.

Browning, C. R., Calder, C. A., Ford, J. L., Boettner, B., Smith, A. L., & Haynie, D. (2017). Understanding racial differences in exposure to violent areas: integrating survey, smartphone, and administrative data resources. *The Annals of the American Academy of Political and Social Science*, *669*(1), 41–62.

Browning, C. R., Calder, C. A., Soller, B., Jackson, A. L., & Dirlam, J. (2017). Ecological networks and neighborhood social organization. *American Journal of Sociology*, *122*(6), 1939–1988.

Connelly, R., Playford, C. J., Gayle, V., & Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research*, *59*(September), 1–12.

Delmelle, E. C. (2016). Mapping the DNA of urban neighborhoods: Clustering longitudinal sequences of neighborhood socioeconomic change. *Annals of the American Association of Geographers*, *106*(1), 36–56.

Desmond, M., Papachristos, A. V., & Kirk, D. S. (2016). Police violence and citizen crime reporting in the black community. *American Sociological Review*, *81*(5), 857–876.

Du Bois, W. E. B. (1996). *The Philadelphia Negro: A social study* (Reprint ed.). Philadelphia: University of Pennsylvania Press.

Glaeser, E. L., Kominers, S. D., Luca, M., & Naik, N. (2018). Big data and big cities: The promises and limitations of improved measures of urban life. *Economic Inquiry*, *56*(1), 114–137.

Golder, S. A., & Macy, M. W. (2014). Digital footprints: Opportunities and challenges for online social research. *Annual Review of Sociology*, *40*(1), 129–152.

Goodspeed, R. (2017). Community and urban places in a digital world. *City & Community*, *16*(1), 9–15.

Hipp, J. R., Butts, C. T., Acton, R., Nagle, N. N., & Boessen, A. (2013). Extrapolative simulation of neighborhood networks based on population spatial distribution: Do they predict crime? *Social Networks*, *35*(4), 614–625.

Hwang, J., & Sampson, R. J. (2014). Divergent pathways of gentrification: Racial inequality and the social order of renewal in Chicago neighborhoods. *American Sociological Review*, *79*(4), 726–751.

Jacobs, J. (1961). *The death and life of great American cities*. New York, NY: Vintage.

Kirk, D. S., & Laub, J. H. (2010). Neighborhood change and crime in the modern metropolis. *Crime and Justice*, *39*(1), 441–502.

Lane, J. (2016). The digital street: An ethnographic study of networked street life in Harlem. *American Behavioral Scientist*, *60*(1), 43–58.

Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., … Gutmann, M. (2009). Life in the network: The coming age of computational social science. *Science (New York, NY)*, *323*(5915), 721.

Legewie, J. (2016). Racial profiling and use of force in police stops: How local events trigger periods of increased discrimination. *American Journal of Sociology*, *122*(2), 379–424.

Legewie, J. (2018). Living on the Edge. Neighborhood Boundaries and the Spatial Dynamics of Violent Crime. *Demography* (forthcoming), 37.

Legewie, J., & Fagan, J. (2018). *Aggressive Policing and the Educational Performance of Minority Youth*. SocArXiv, 29 Aug. 2018. https://doi.org/10.31235/osf.io/rdchf.

Legewie, J., & Schaeffer, M. (2016). Contested boundaries: Explaining where ethnoracial diversity provokes neighborhood conflict. *American Journal of Sociology*, *122*(1), 125–161.

Minkoff, S. L. (2016). NYC 311: A tract-level analysis of citizen–government contacting in New York City. *Urban Affairs Review*, *52*(2), 211–246.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., … Yarkoni, T. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422.

O'Brien, D. T., Sampson, R. J., & Winship, C. (2015). Ecometrics in the age of big data: Measuring and assessing "broken windows" using large-scale administrative records. *Sociological Methodology*, *45*(1), 101–147.

Papachristos, A. V., Hureau, D. M., & Braga, A. A. (2013). The corner and the crew: The influence of geography and social networks on gang violence. *American Sociological Review*, *78*(3), 417–447.

Salganik, M. J. (2017). *Bit by bit: Social research in the digital age*. Princeton, NJ: Princeton University Press.

Sampson, R. J. (2012). *Great American city: Chicago and the enduring neighborhood effect*. Chicago, IL: University of Chicago Press.

Small, M. L. (2017). *Someone to talk to*. Oxford: Oxford University Press.

Smith, C. M., & Papachristos, A. V. (2016). Trust thy crooked neighbor: Multiplexity in Chicago organized crime networks. *American Sociological Review*, *81*(4), 644–667.

Tina Law (M.A., Yale University) is a PhD student in sociology at Northwestern University. Her research focuses on urban sociology, racial inequality, neighborhoods, social networks, and computational social science. In

particular, her research explores how administrative and archival "big data" and computational social science methods can be used to advance the study of urban change and racial inequality in the United States. She is a National Science Foundation Graduate Research Fellow.

Joscha Legewie (PhD 2013, Columbia University) is an assistant professor of sociology at Harvard University. His research focuses on social inequality/stratification, race/ethnicity, quantitative methods, education, urban sociology, and computational social science. His work is based on innovative quantitative methods. It builds on rigorous causal inference using natural or quasi-experimental research designs with a keen interest in "big data" as a promising source for future social science research—including administrative student records, millions of time and geo-coded NYPD stop-and-frisk operations or 311 service requests from New York City. His research was published in *the American Journal of Sociology*, *the American Sociological Review*, *Sociology of Education* and other major journals.

RELATED ESSAYS

Cities and Sustainable Development *(Sociology)*, Christopher Cusack
Neighborhoods and Cognitive Development *(Psychology)*, Jondou Chen and Jeanne Brooks-Gunn
Sociological Theory After the End of Nature *(Sociology)*, Robert J. Brulle
Theorizing the Death of Cities *(Political Science)*, Peter Eisinger
Exploring Opportunities in Cultural Diversity *(Political Science)*, David D. Laitin and Sangick Jeon
Organizational Populations and Fields *(Sociology)*, Heather A. Haveman and Daniel N. Kluttz
Digital Methods for Web Research *(Methods)*, Richard Rogers
Longitudinal Data Analysis *(Methods)*, Todd D. Little *et al*.
An Emerging Trend: Is Big Data the End of Theory? *(Sociology)*, Michael W. Macy