# 5. Big data and the city

*Matthew Zook, Taylor Shelton and Ate Poorthuis*

## 5.1 CITIES AND THE RISE OF BIG DATA

As more and more aspects of contemporary urban society are tracked and quantified, the emerging cloud of so-called 'big data' is widely considered to represent a fundamental change in the way we interact with and understand cities. For some proponents of big data, like Anderson (2008), big data means the 'end of theory' and the ability to let 'the numbers speak for themselves'. For others, big data allows for the development of a single, ostensibly 'universal' theory of the urban that moves urban studies towards a greater standing in the scientific community (Bettencourt and West, 2010; Lehrer, 2010). Regardless of the particular context, the prevailing discourse around big data in the first half of the 2010s has been founded on the 'widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy' (Boyd and Crawford, 2012, p. 663).

While these emerging data-driven understandings of cities often run counter to a more theoretical and heterodox approach to urban geography (Wyly, 2014), it is worth noting that these trends also pre-date the emergence of what we now call 'big data'. For over a century urban planners, policymakers and social scientists have pursued a data-based, scientific and technologically-mediated understanding of urban life (cf. Ford, 1913; Fairfield, 1994; Light, 2003; LeGates et al., 2009; Barnes, 2013; Barnes and Wilson, 2014). While much of the initial work on big urban data has focused on the novelty of such datasets, the challenge moving forward lies in combining these new data sources and computational methods with established theoretical approaches applied to longstanding questions, as in the case of recent work by Cranshaw et al. (2012), Arribas-Bel (2015) and Shelton et al. (2015) utilizing social media data to understand the dynamics of urban segregation, mobility and neighbourhood change.

In order to illustrate the potential of big data for urban geographic research, we explore how these data sources and methods might be usefully applied to the persistent question of gentrification. We first review how gentrification has been defined and measured in the existing literature, and how these definitions and metrics have shaped our understandings of the process. Next, we outline nascent attempts to use big data, especially social media data, to understand gentrification. We pay attention to more 'naïve' approaches that draw upon big data but in ways that do not fully engage with its messy and complicated nature, or which fail to connect with longer standing approaches within urban geography. We then contrast these perspectives with a range of more constructive possibilities for using big data to study gentrification that build from existing scholarship and recognize both the advantages and disadvantages of big data over other more conventional forms of data used in previous research. In short, we argue that big data is unlikely to be a panacea for empirical studies of gentrification, or for any particular urban issue of interest, and the 'multidimensionality of gentrification' still means that 'the use of a single variable to identify it is almost certain to fail' (Bostic and Martin, 2003, p. 2431). We do argue, however, that big data can supplement existing data sources and provide a richer understanding of the multiple social and spatial processes that characterize the process of gentrification, its constituent parts, causes and effects.

## 5.2 CONVENTIONAL APPROACHES TO STUDYING GENTRIFICATION

Gentrification – broadly defined as a transition in neighbourhood character associated with the displacement of lower-income residents by higher-income newcomers (Glass, 1964) – has been studied extensively for the last half-century. But as gentrification has accelerated in recent years, it has garnered newfound attention, especially in the mainstream press, leading some popular interpretations to seize upon scholarly debates about gentrification as a 'chaotic concept' (Rose, 1984) and question its utility, or even reality, as an approach to studying cities (Buntin, 2015; *The Economist*, 2015; Cortright, 2015). In part, this argument over whether gentrification exists is tied to the differences in how gentrification is defined and measured by assorted scholars. As Hammel and Wyly (1996, p. 248) argue, '[t]he uncertainty over the extent of gentrification stems not only from the complexity of the process, but also from the difficulty of observing and measuring the phenomenon'. At the same time, however, this 'infatuation with how to define the process' (Slater, 2006, p. 744) risks blunting the critical tenor driving much gentrification research. It is thus necessary to strike a balance between the identification of good metrics for gentrification research while not allowing such metrics to become the ends, in and of themselves.

By and large, most conventional analyses of gentrification use census data and other state-sourced secondary data to track change over time in key indicators, such as: household income, educational attainment, property values and rents, and less often, racial composition. While changes in an area's median household income represent the most straightforward indicator of changing class structure, focusing only on income can be problematic, ignoring differences between 'gentrifiable' and 'non-gentrifiable' areas, and failing to recognize that changing class composition is not always reflected most directly in income. To capture these more complex class dynamics, alternative indicators might be better proxies for gentrification, such as the proportion of residents with a college education (Schuler et al., 1992; Hammel and Wyly, 1996; Freeman, 2005) or in professional or managerial careers (Atkinson, 2000). The use of race or ethnicity as an indicator of gentrifying neighbourhoods has been less commonly used, given that most analyses focus on gentrification as an essentially class-based process. But as many gentrifying neighbourhoods, especially in the United States, have higher than average proportions of minority (especially black) residents, race remains an inextricable aspect of how gentrification is experienced and interpreted in many places (Kirkland, 2008).

Gentrification is also evident in material changes to neighbourhoods' built environments and thus measures such as mortgage borrowing (Wyly and Hammel, 1999; Kreager et al., 2011) and owner-occupied housing (Heidkamp and Lucas, 2006) have served as metrics as well. Foundational work by Wyly and Hammel (1998, 1999; Hammel and Wyly, 1996) provides one of the most comprehensive assessments of these indicators, combining census data with an inductive, field-based survey of building conditions and upgrading activity to demonstrate how trends in these indicators align with different stages in the process of neighbourhood change. While not using big data per se, aspects of this approach have been updated in the digital era, as with Hwang and Sampson's (2014) use of Google Street View to carry out a large-scale digital windshield survey of changes in neighbourhood built environments.

Despite a general recognition that '[d]isplacement is vital to an understanding of gentrification, in terms both of retaining definitional coherence and of retaining a critical perspective on the process' (Slater et al., 2004, p. 1144), it remains arguably the most difficult aspect of the process to measure. Census-type data showing loss of low-income or racial/ethnic-minority households and gains in more affluent and white households can infer displacement, but capturing the why of particular shifts remains elusive. Like the aforementioned attempts to capture on-the-ground upgrading processes, detailed descriptions and explanations of displacement are more complex and resource-intensive processes than most analyses are capable of grappling with. Further complicating this picture are attempts to destabilize the conventional interpretation of gentrification as negative for existing residents. For instance, work by Freeman (2005; Freeman and Braconi, 2004) and Vigdor (2002) has argued that residents of gentrifying neighbourhoods are in fact less likely to move or be displaced than the average resident of a non-gentrifying neighbourhood.

More recently, researchers have also used less conventional, non-official (albeit not 'big') sources of data to measure gentrification. In addition to indicators of the built environment referenced above, scholars have used localized knowledge such as newspapers, non-profits and community groups to identify gentrifying neighbourhoods. These definitions often employ different criteria and as a result can vary significantly from academic studies, as Barton's (2016) comparison of gentrifying neighbourhoods identified by *The New York Times* with those of Bostic and Martin (2003) and Freeman (2005) demonstrates. Other studies have focused on specific kinds of activities or businesses – such as coffee or cupcake shops (Papachristos et al., 2011; Smith, 2014; Twilley, 2009, 2011) – seen as indicative of gentrification's impact on consumptive activities and commercial spaces. While these kinds of non-residential indicators can 'provide an on-the-ground and visible manifestation of a particular form of gentrification – the increased presence of an amenity often associated with gentrifiers' lifestyles' (Papachristos et al., 2011, p. 216), this approach is difficult to replicate over time; cupcake shops might be indicative of gentrification now, but cultural tastes evolve quickly and this kind of indicator is unlikely to be useful for analyses in the more distant past or future.

## 5.3 POSSIBLE BIG DATA APPROACHES TO GENTRIFICATION

Together, these competing understandings of gentrification and approaches to measuring it lead to questions of how emerging sources of big data might be used to add something new to our understanding of this process. In particular, one of the biggest advantages big data offers is the ability to overcome the persistent limitations of spatial and temporal scale in conventional data that are often released at large or irregular time intervals and aggregated to relatively coarse administrative geographies like census tracts (Schuler et al., 1992) divorced from the way neighbourhood change actually unfolds. Some studies – such as Freeman (2005) and Freeman and Braconi (2004) in New York, Vigdor (2002) in Boston and Papachristos et al. (2011) in Chicago – even use neighbourhood definitions that are substantially larger than census tracts, with some 'neighbourhoods' including over 100,000 people within their boundaries. This is clearly problematic as gentrification occurs more rapidly than a decennial census can detect, and unfolds over irregular geographies, along commercial corridors or other points of interest, which may include parts of multiple administrative or statistical geographies, while encompassing none in their entirety. Indeed, the process of gentrification often results in the redefinition of neighbourhood names and boundaries as the character of previously distinct places evolves (Madden, 2017). New sources of big data, on the other hand, are often produced and collected continuously in real-time, while also being point-based, allowing for aggregation to multiple spatial scales, making them particularly relevant to tracking these finer-grained changes in the urban fabric. That being said, these more fine-grained sources of data have only emerged relatively recently so might not yet hold utility for understanding changes over longer periods of time, which still requires augmentation with some more conventional sources of social data.

Despite these advantages, much of the work to date that uses big data to study gentrification has been narrowly focused and primarily descriptive, more about novelty than a substantively improved understanding of the process. For example, Beekmans (2011) uses Foursquare check-ins in Amsterdam to examine gentrifying neighbourhoods. Although this work was certainly novel and commendable in 2011, it remains primarily a descriptive mapping of Foursquare venues and check-in densities accompanied by a visual comparison of locations deemed to be gentrifying. Looking at the more negative aspects of gentrification, Schaefer (2014) and Venerandi et al. (2015) use tweets and Foursquare and OpenStreetMap data, respectively. Similar to Beekmans, they use their data roughly as point-of-interest data: the more points related to displacement or deprivation, the higher the 'score' for that location. Although these approaches provide a different view on gentrification than more conventional census indicators, it replaces one imperfect measure with another. Ultimately, the strengths of representative sampling and other safeguards against bias and inaccuracies might make census data a more reliable data source for rigorous analysis. And while Schaefer (2014) also attempts to analyse keyword frequencies in this data – including words such as 'loft', 'eviction', 'yuppie' and 'gentrification' – in his case study of Los Angeles, the few thousand tweets matching these criteria suggest that reliance on Twitter keywords alone may not be enough to detect the occurrence of gentrification (or conversations about it).

Although some of this early work might indeed be labelled naïve in its conceptualization and operationalization of gentrification, we can identify more informed approaches in some more recent studies. For example, Hristova et al. (2016) attempt to move beyond mere description by harnessing a dataset derived from both Twitter and Foursquare activity in London. They not only take into account check-ins at locations, but also consider at which categories of establishments these check-ins are made, constructing a kind of spatial profile of each person checking in that includes the user's social ties and what other types of locations they have checked into. In this way, they construct a set of diversity indicators for each location, differentiating between those places that represent meeting points for friends or strangers, the diversity of visits and visitors to a given neighbourhood and the likelihood of serendipitous encounters. Comparing these various diversity metrics with census-derived deprivation indicators allows them to suggest that neighbourhoods with both high diversity and deprivation might overlap with neighbourhoods commonly thought of as gentrifying. Ultimately, these social media-based diversity indicators provide an example of going beyond the aforementioned descriptive approaches, capturing a more varied set of social and spatial processes through big data that would not be available through more conventional data sources, while also showing the potential for combining these new approaches with more conventional indicators to produce a synthetic analysis of gentrification.

To address these shortcomings and build on the potential shown by Hristova et al. (2016), researchers could use a larger variety of big data sources either independently or in combination to understand the process of gentrification in a more holistic and multidimensional manner. In addition to questions of spatial and temporal scale, one of the key limiting factors of conventional spatial data sources is that they tie a person to a single point in space, most often where they sleep. However, people are fundamentally mobile, moving between and within many different locations, even in the course of a single day. In other words, it is highly limiting to sum up the very complex social lives of individuals, occurring across and within so many different places, into a single, fixed residential location (Kwan, 2012, 2013). This is especially relevant for gentrification research, as early phases of the gentrification process might not be linked solely to people moving to a new neighbourhood to take up residence, but also people moving through a neighbourhood to eat and drink, to go to art and music shows, or simply to hang out with friends who live there.

While the finer spatial and temporal resolution of big data allows for some greater degree of accuracy in analysing gentrification, the ability to capture these everyday mobilities also provides an opportunity to rethink the underlying geographies of the city and how gentrification is a process that itself reshapes urban geographies both materially and imaginatively. Rather than relying on, and reifying, conventional administrative or statistical geographies, such data can be used to demonstrate the fundamental connectedness of different neighbourhoods within the city, and how these connections are constantly changing. Indeed, big data now allows us to study these types of interconnections both on a daily scale, as well as via longer-term, multi-year longitudinal studies. There is already a wide variety of existing work on this type of urban mobility that can potentially be extended to look at gentrification processes specifically. Data sources that enable such analyses include social media data (Shelton et al., 2015), taxi and ride-share data (Liu et al., 2012), transit smart card data (Hasan et al., 2013), and mobile phone data (Järv et al., 2014, 2015). All of these data sources provide potential insights into daily travel and interaction patterns, allowing researchers to understand the spatiality of gentrification beyond simply the location of individual residences.

Big data can also provide a window into other aspects of gentrification that are not well captured in conventional data sources. For example, with respect to the housing market itself, data from online platforms such as Craigslist, Zillow and Zoopla offer a supplement to more conventional data sources of property transactions or valuations, which typically fail to capture the crucially important role of the rental market in gentrification. While still relatively unexploited within the literature, recent work by Boeing and Waddell (2016) using Craigslist data, and by Wachsmuth (2017) on Airbnb short-term rental listings, suggests the potential for these platforms to provide insight on gentrification.

Big data can similarly provide more fine-grained insights into the place of consumptive behaviour within the gentrification process. As mentioned before, researchers have already explored the role of specific kinds of consumption practices associated with gentrification, such as the sudden popularity of coffee shops and cupcake bakeries in a neighbourhood. But understanding these changes on a larger scale can be difficult, as even the *North American Industry Classification* system, with over a thousand industry classes, is unable to distinguish between a Dunkin Donuts chain or a shop that sells handcrafted, artisanal 'cronuts'. Through the myriad of online platforms that people use to publicize their consumptive behaviour – from specialized sites like Yelp to more general purpose apps Foursquare and Facebook – we can not only attempt to answer the question 'who consumes what where?', but also subsequent questions such as 'where else do people visiting a restaurant with a certain profile go?'. Some early work already shows promise in this direction: Foursquare check-ins can be used to look at the spatial and 'cultural' footprints of cities and neighbourhoods (Silva et al., 2013, 2014), but can also be used to analyze socio-economic hierarchical structures within cities (Fekete, 2014; van Meeteren and Poorthuis, 2017).

Some of these same processes might also be captured in non-social media sources of big data. For instance, data on credit card transactions or from mobile payment systems, which contains information on the consumer's home location, consumption location, type of consumption and amount of money spent (de Montjoye et al., 2015; Singh et al., 2015), could provide insight into both conventional retail consumption as well as in non-brick-and-mortar staples in gentrifying spaces, such as farmers' markets, food trucks and pop-up shops. Aggregated data from financial companies and credit bureaus, long used within geodemographic profiling (Singleton and Spielman, 2014), could even provide insight into the kind of individual or neighbourhood-level financial distress that helps to precipitate gentrification and displacement.

Despite all of the potential applications of big data to studying gentrification and other urban geographic processes, these approaches and data sources represent only one small part of the available data, namely the part that is easily quantified and machine-readable. That is, each of these datasets and their utility is premised on the presence of locational data, time stamps and a limited set of quantitative variables and categories that serve as proxies for different kinds of socio-economic behaviour. However, many big data sources also contain rich qualitative data, ranging from restaurant reviews on Yelp, to tweets and Facebooks posts, to photos shared on Instagram and Flickr (cf. Zukin et al., 2015 for an analysis of the connections between restaurant reviews on Yelp and gentrification). This type of data is much more challenging to analyse in a comprehensive manner, but geographers are particularly well-equipped to marry quantitative and qualitative approaches to big data (DeLyser and Sui, 2013), which hold significant potential for studying the nuances of gentrification processes and how they are experienced and talked about by people in these places. On the one hand, qualitative methods are necessary to understand and analyse this data in all of its complexity, but are insufficiently equipped to handle its large volume (Jung, 2015). On the other hand, quantitative analysis, with approaches to textual and visual analysis such as machine learning, can process much larger volumes of data but can do so, for now, only in fairly crude ways (cf. Naik et al., 2014 and Liu et al., 2016 on the use of Google Street View to assess perceptions of the built environment). Although very promising, the veritable cutting edge of the field, this kind of hybrid approach has yet to be developed systematically in a way that convincingly demonstrates its potential for understanding gentrification and other urban processes.

Ultimately, despite the greater spatial and temporal granularity in these data and the creative potential for combining quantitative and qualitative analysis, challenges to using big data sources in urban geographic research continue to exist. Perhaps the most significant barrier is the difficulty of gaining access to the data in the first place. While national statistics bureaus often make data available to all researchers, many of the private entities that produce or facilitate the creation of big data might not make their data accessible at all. Or, as the case often is, only researchers with a significant degree of technical acumen or social capital are capable of accessing and analysing the data, reproducing a hierarchy between different social science methods (Poorthuis and Zook, 2017). Apart from issues of data access, there are also lingering issues around the validity and reliability of big data. How big data is collected and constructed differs radically from conventional sampling-based research and each big data source might introduce its own (potentially unknown) issues around who is represented within the dataset and what biases are introduced. On top of all this, there are also significant issues around privacy and the ethics of doing research with such datasets, with no scholarly consensus on how these issues should best be addressed (cf. Boyd and Crawford, 2012; Kitchin, 2013, 2014; Zook et al., 2017).

Even were one to sufficiently address each of these issues, there would remain the question of whether big data is actually suitable for capturing the fundamental processes that underlie gentrification. As has been shown in a number of other contexts, the 'bigness' of the data does not necessarily make it appropriate for the given question at hand (Lazer et al., 2014). Tweets or Instagram photos are probably not the best means of quantifying displacement or the rent gap, among other difficult-to-quantify concepts. Though social media data is incredibly useful for understanding any number of social and spatial processes, we have to make sure that the indicators we gather from social media are indeed good indicators or proxies for the social process we are interested in, whether it be gentrification or anything else.

## 5.4 CONCLUSION

Moving forward, the imperatives for urban geographical research using big data are two-fold. First, for those more computationally-minded scholars looking to apply these data sources to urban questions, it is necessary to avoid the naiveté that would lead one to assume that this data can provide a substantive understanding of the world without simultaneously being grounded in the requisite theoretical perspectives to inform such an analysis. Knowing which questions to ask, and how those questions might be approximated by the data at hand, are crucial to any big data research. Second, for critical urban geographers, it is important not to discount the potential of big data simply because it is often used in these more naïve ways. Rather than being an inherent feature of the data, this is very much a result of the particular contingent circumstances that bring together data, methodology, epistemology and politics to produce these approaches (cf. Wyly, 2009).

Gentrification, like other urban social and spatial processes, is so variegated and multifaceted that any single definition, let alone any single quantitative indicator, cannot do it justice. We should not expect big data to alter this dynamic. But even if big data cannot provide a single, universal way of understanding gentrification, it can offer alternative ways of analysing this and other sociospatial processes that complement and extend, rather than replace, existing methods and approaches. Indeed, it is through the creative combination of these data sources and methods that we might be able to extend our empirical and theoretical understanding of urban geography in all of its multidimensionality.

## REFERENCES

Anderson, C. (2008), 'The end of theory: the data deluge makes the scientific method obsolete', *Wired Magazine*, 23 June 2008, at: https://www.wired.com/2008/06/pb-theory/ (accessed 18 September 2018).

Arribas-Bel, D. (2015), 'The spoken postcodes', *Regional Studies, Regional Science*, **2** (1), 458–461.

Atkinson, R. (2000), 'Measuring gentrification and displacement in Greater London', *Urban Studies*, **37** (1), 149–165.

Barnes, T. J. (2013), 'Big data, little history', *Dialogues in Human Geography*, **3** (3), 297–302.

Barnes, T. J. and M. W. Wilson (2014), 'Big data, social physics, and spatial analysis: the early years', *Big Data & Society*, **1** (1), 1–14.

Barton, M. (2016), 'An exploration of the importance of the strategy used to identify gentrification', *Urban Studies*, **53** (1), 92–111.

Beekmans, J. (2011), 'Check-in urbanism: exploring gentrification through Foursquare activity', MSc dissertation, University of Amsterdam, at: http://www.scriptiesonline.uba.uva.nl/391330 (accessed 18 September 2018).

Bettencourt, L. and G. West (2010), 'A unified theory of urban living', *Nature*, **467** (7318), 912–913.

Boeing, G. and P. Waddell (2016),'New insights into rental housing markets across the United States: web scraping and analyzing Craigslist rental listings', *Journal of Planning Education and Research*, **37** (4), 457–476.

Bostic, R. W. and R. W. Martin (2003), 'Black home-owners as a gentrifying force? Neighbourhood dynamics in the context of minority home-ownership', *Urban Studies*, **40** (12), 2427–2449.

Boyd, D. and K. Crawford (2012), 'Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon', *Information, Communication & Society*, **15** (5), 662–679.

Buntin, J. (2015), 'The myth of gentrification', *Slate*. 14 January, at: http://www.slate.com/articles/news_and_politics/politics/2015/01/the_gentrification_myth_it_s_rare_and_not_as_bad_for_the_poor_as_people.html (accessed 18 September 2018).

Cortright, J. (2015), 'In defense of gentrification', *The Atlantic*, 31 October, at: http://www.theatlantic.com/business/archive/2015/10/in-defense-of-gentrification/413425/ (accessed 18 September 2018).

Cranshaw, J., R. Schwartz, J. I. Hong, and N. Sadeh (2012), 'The Livehoods Project: utilizing social media to understand the dynamics of a city', *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*, 58–65.

DeLyser, D. and D. Sui (2013), 'Crossing the qualitative-quantitative divide II: inventive approaches to big data, mobile methods, and rhythm analysis', *Progress in Human Geography*, **37** (2), 293–305.

de Montjoye, Y., L. Radaelli, V. K. Singh, and A. Pentland (2015), 'Unique in the shopping mall: on the reidentifiability of credit card metadata', *Science*, **347** (6221), 536–539.

Fairfield, J. D. (1994), 'The scientific management of urban space: professional city planning and the legacy of progressive reform', *Journal of Urban History*, **20** (2), 179–204.

Fekete, E. (2014), 'Consumption and the urban hierarchy in the southeastern United States', *Southeastern Geographer*, **54** (3), 249–269.

Ford, G. (1913), 'The city scientific', *Engineering Record*, **67**, 551–552.

Freeman, L. (2005), 'Displacement or succession? Residential mobility in gentrifying neighborhoods', *Urban Affairs Review*, **40** (4), 463–491.

Freeman, L. and F. Braconi (2004), 'Gentrification and displacement New York City in the 1990s', *Journal of the American Planning Association*, **70** (1), 39–52.

Glass, R. (1964), *London: Aspects of Change*, London: MacGibbon & Kee.

Hammel, D. J. and E. K. Wyly (1996), 'A model for identifying gentrified areas with census data', *Urban Geography*, **17** (3), 248–268.

Hasan, S., C. M. Schneider, S. V. Ukkusuri, and M. C. González (2013), 'Spatiotemporal patterns of urban human mobility', *Journal of Statistical Physics*, **151** (1–2), 304–318.

Heidkamp, C. P. and S. Lucas (2006), 'Finding the gentrification frontier using census data: the case of Portland, Maine', *Urban Geography*, **27** (2), 101–125.

Hristova, D., M. J. Williams, M. Musolesi, P. Panzarasa and C. Mascolo (2016), 'Measuring urban social diversity using interconnected geo-social networks', *Proceedings of the 25th International Conference on World Wide Web*, 21–30.

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

Järv, O., R. Ahas and F. Witlox (2014), 'Understanding monthly variability in human activity spaces: a twelve-month study using mobile phone call detail records', *Transportation Research Part C: Emerging Technologies*, **38**, 122–135.

Järv, O., K. Müürisepp, R. Ahas, B. Derudder and F. Witlox (2015), 'Ethnic differences in activity spaces as a characteristic of segregation: a study based on mobile phone usage in Tallinn, Estonia', *Urban Studies*, **52** (14), 2680–2698.

Jung, J. (2015), 'Code clouds: qualitative geovisualization of geotweets', *The Canadian Geographer / Le Géographe Canadien*, **59** (1), 52–68.

Kirkland, E. (2008), 'What's race got to do with it? Looking for the racial dimensions of gentrification', *Western Journal of Black Studies*, **32** (2), 18–30.

Kitchin, R. (2013), 'Big data and human geography: opportunities, challenges and risks', *Dialogues in Human Geography*, **3** (3), 262–267.

Kitchin, R. (2014), 'Big data, new epistemologies and paradigm shifts', *Big Data & Society*, **1** (1), 1–12.

Kreager, D. A., C. J. Lyons and Z. R. Hays (2011), 'Urban revitalization and Seattle crime, 1982–2000', *Social Problems*, **58** (4), 615–639.

Kwan, M. (2012), 'The uncertain geographic context problem', *Annals of the Association of American Geographers*, **102** (5), 958–968.

Kwan, M. (2013), 'Beyond space (as we knew it): toward temporally integrated geographies of

segregation, health, and accessibility', *Annals of the Association of American Geographers*, **103** (5), 1078–1086.
Lazer, D., R. Kennedy, G. King and A. Vespignani (2014), 'The parable of Google flu: traps in big data analysis', *Science*, **343** (6176), 1203–1205.
LeGates, R., N. J. Tate and R. Kingston (2009), 'Spatial thinking and scientific urban planning', *Environment and Planning B: Planning and Design*, **36** (5), 763–768.
Lehrer, J. (2010), 'A physicist turns the city into an equation', *The New York Times*, 17 December, at: http://www.nytimes.com/2010/12/19/magazine/19Urban_West-t.html (accessed 18 September 2018).
Light, J. S. (2003), *From Warfare to Welfare: Defense Intellectuals and Urban Problems in Cold War America*, Baltimore, MA: Johns Hopkins University Press.
Liu, Y., C. Kang, S. Gao, Y. Xiao and Y. Tian (2012), 'Understanding intra-urban trip patterns from taxi trajectory data', *Journal of Geographical Systems*, **14** (4), 463–483.
Liu, L., H. Wang and C. Wu (2016), 'A machine learning method for the large-scale evaluation of urban visual environment', arXiv Preprint: 1608.03396, at: https://arxiv.org/abs/1608.03396 (accessed 19 September 2018).
Madden, D. J. (2017), 'Pushed off the map: toponymy and the politics of place in New York City', *Urban Studies*, **55** (8), 1599–1614.
Naik, N., R. Raskar and C. Hidalgo (2014), 'Streetscore – predicting the perceived safety of one million streetscapes', *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 779–785.
Papachristos, A. V., C. M. Smith, M. L. Scherer and M. A. Fugiero (2011), 'More coffee, less crime? The relationship between gentrification and neighborhood crime rates in Chicago, 1991 to 2005', *City & Community*, **10** (3), 215–240.
Poorthuis, A. and M. Zook (2017), 'Making big data small: strategies to expand urban and geographical research using social media', *Journal of Urban Technology*, **24** (4), 115–135.
Rose, D. (1984), 'Rethinking gentrification: beyond the uneven development of Marxist urban theory', *Environment and Planning D: Society and Space*, **2** (1), 47–74.
Schaefer, B. (2014), 'Social media to locate urban displacement: assessing the risk of displacement using volunteered geographic information in the city of Los Angeles', M.Sc. dissertation, University of Southern California, at: http://digitallibrary.usc.edu/cdm/ref/collection/p15799coll3/id/514559 (accessed 18 September 2018).
Schuler, J. R., R. B. Kent and C. B. Monroe (1992), 'Neighborhood gentrification: a discriminant analysis of a historic district in Cleveland, Ohio', *Urban Geography*, **13** (1), 49–67.
Shelton, T., A. Poorthuis and M. Zook (2015), 'Social media and the city: rethinking urban socio-spatial inequality using user-generated geographic information', *Landscape and Urban Planning*, **142**, 198–211.
Silva, T. H., P. O. S. Vaz de Melo, J. Almeida, M. Musolesi and A. Loureiro (2014), 'You are what you eat (and drink): identifying cultural boundaries by analyzing food and drink habits in Foursquare', Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media, at https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewFile/8113/8147 (last accessed 19 September 2018).
Silva, T. H., P. O. S. Vaz de Melo, J. M. Almeida, J. Salles and A. A. F. Loureiro (2013), 'A comparison of Foursquare and Instagram to the study of city dynamics and urban social behavior', *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*, no pagination.
Singh, V. K., B. Bozkaya and A. Pentland (2015), 'Money walks: implicit mobility behavior and financial well-being', *PloS One*, **10** (8), e0136628.
Singleton, A. D. and S. E. Spielman (2014), 'The past, present, and future of geodemographic research in the United States and United Kingdom', *The Professional Geographer*, **66** (4), 558–567.
Slater, T. (2006), 'The eviction of critical perspectives from gentrification research', *International Journal of Urban and Regional Research*, **30** (4), 737–757.
Slater, T., W. Curran and L. Lees (2004), 'Gentrification research: new directions and critical scholarship', *Environment and Planning A*, **36** (7), 1141–1150.

Smith, C. M. (2014), 'The influence of gentrification on gang homicides in Chicago neighborhoods, 1994 to 2005', *Crime & Delinquency*, **60** (4), 569–591.
*The Economist* (2015), 'Bring on the hipsters: gentrification is good for the poor', *The Economist*, 19 February, at: https://www.economist.com/news/united-states/21644164-gentrification-good-poor-bring-hipsters (accessed 18 September 2018).
Twilley, N. (2009), 'Cupcake gentrification', *Edible Geography Blog*, 5 August, at: http://www.ediblegeography.com/cupcakegentrification/ (accessed 19 September 2018).
Twilley, N. (2011), 'Mapping gangs and cupcakes', *Edible Geography Blog*, 9 March, at: http://www.ediblegeography.com/mapping-gangs-and-cupcakes, (accessed 19 September 2018).
van Meeteren, M. and A. Poorthuis (2017), 'Christaller and "big data": recalibrating central place theory via the geoweb', *Urban Geography*, **39** (1), 122–148.
Venerandi, A., G. Quattrone, L. Capra, D. Quercia and D. Saez-Trumper (2015), 'Measuring urban deprivation from user generated content', *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 254–264.
Vigdor, J. L. (2002), 'Does gentrification harm the poor?' *Brookings-Wharton Papers on Urban Affairs*, 133–182.
Wachsmuth, D. (2017), 'Airbnb and gentrification in New York', *David Wachsmuth personal blog*, 13 March, at: https://davidwachsmuth.com/2017/03/13/airbnb-and-gentrification-in-new-york/ (accessed 18 September 2018).
Wyly, E. K. (2009), 'Strategic positivism', *The Professional Geographer*, **61** (3), 310–322.
Wyly, E. K. (2014),'Automated (post)positivism', *Urban Geography*, **35** (5), 669–690.
Wyly, E. K. and D. J. Hammel (1998), 'Modeling the context and contingency of gentrification', *Journal of Urban Affairs*, **20** (3), 303–326.
Wyly, E. K. and D. J. Hammel (1999), 'Islands of decay in seas of renewal: housing policy and the resurgence of gentrification', *Housing Policy Debate*, **10** (4), 711–771.
Zook, M., S. Barocas, D. Boyd, K. Crawford, E. Keller, S. P. Gangadharan, A. Goodman, R. Hollander, B. A. Koenig, J. Metcalf, A. Narayanan, A. Nelson and F. Pasquale (2017), 'Ten simple rules for responsible big data research', *PLOS Computational Biology*, **13** (3), e1005399.
Zukin, S., S. Lindeman, and L. Hurson (2015), 'The omnivore's neighborhood? Online restaurant reviews, race, and gentrification', *Journal of Consumer Culture*, **17** (3), 459–479.