Article

Urban Analytics and City Science

EPB: Urban Analytics and City Science
2019, Vol. 46(9) 1756–1768

DOI: 10.1177/2399808319882826
journals.sagepub.com/home/epb
SAGE

A roundtable discussion: Defining urban data science

Organizers

Wei Kang
University of California, Riverside, USA

Taylor Oshan
University of Maryland, USA

Levi J Wolf
University of Bristol, UK

Discussants

Geoff Boeing
University of Southern California, USA

Vanessa Frias-Martinez
University of Maryland Institute for Advanced Computer Studies, USA

Song Gao
University of Wisconsin, Madison, USA

Ate Poorthuis
Singapore University of Technology and Design, Singapore

Wenfei Xu
Columbia University, USA

Abstract

The field of urban analytics and city science has seen significant growth and development in the past 20 years. The rise of data science, both in industry and academia, has put new pressures on

Corresponding author:
Levi J Wolf, School of Geographical Sciences, University of Bristol, Bristol, UK.
Email: levi.john.wolf@bristol.ac.uk

urban research, but has also allowed for new analytical possibilities. Because of the rapid growth and change in the field, terminology in urban analytics can be vague and unclear. This paper, an abridged synthesis of a panel discussion among scholars in Urban Data Science held at the 2019 American Association of Geographers Conference in Washington, D.C., outlines one discussion seeking a better sense of the conceptual, terminological, social, and ethical challenges faced by researchers in this emergent field. The panel outlines the difficulties of defining what is or is not urban data science, finding that good urban data science must have an expansive role in a successful discipline of "city science." It suggests that "data science" has value as a "signaling" term in industrial or popular science applications, but which may not necessarily be well-understood within purely academic circles. The panel also discusses the normative value of doing urban data science, linking successful practice back to urban life. Overall, this panel report contributes to the wider discussion around urban analytics and city science and about the role of data science in this domain.

Keywords
City analytics, urban data

Introduction

As is frequently discussed in the literature on urban analytics and city science, the clarification and maintenance of terminologies is important for the domain (Batty, 2019a, 2019b). Indeed, understanding what a city is or is not (Beynon et al., 2016; Deavers, 1992; Hall et al., 2006; Hoggart, 1988; Isserman, 2005; Krupat and Guild, 1980; McDade and Adair, 2001; Pryor, 1968; Wilson, 1968) precludes much of the contemporary confusion. Attempting to understand whether a city is "smart" (Breuer et al., 2014; Kitchin, 2016; Nam and Pardo, 2011; Thompson, 2016), where it stops and starts, begins and ends (Borruso, 2003; Small et al., 2005; Tannier et al., 2011; Taylor, 2011; Tsai, 2005; Yin et al., 2017), or the form of its internal physical, social, or biological structure (Boeing, 2019; Chodrow, 2017; Evans et al., 2011; Frias-Martinez et al., 2012; Gao et al., 2013; Louail et al., 2014; Naik et al., 2016; O'Brien et al., 2015; Shelton and Poorthuis, 2019; Williams et al., 2019) all feel premature without a solid understanding of the scope, depth, and structure of "urban analytics." However, "the massive confusion of language and terminology" has become "a continuing feature of our field" (Batty, 2019b: 997). This confusion makes it challenging for urban analytics and city science to be communicated clearly and impairs cooperation with other sciences.

However, this unclear language impairs how we understand our own discipline as well. A few recent attempts to define the field of "urban analytics" or "city analytics" have focused on defining the discipline and outlining the set of activities it entails. For instance, Higham et al. (2017) suggest that "city analytics" is different from previous studies of the city because it is both "challenging" and "interdisciplinary ... interact[ing] with multiple sectors ... and key professions" (1). It "involves disparate new types of data" and "raises novel issues in terms of privacy and ethics," (2) because of the structure and pervasiveness of these new data. This "new"ness in urban analytics is noted by Singleton et al. (2017) as well, since "urban analytics is the practice of using new forms of data in combination with computational approaches to gain insight into urban processes" (15) and Batty (2019a) notes Goodchild's definition as a "new kind of urban research ... that exploits the vast new data resources that are becoming available." Singleton and Arribas-Bel (2019) suggest that "geographic data science," closely linked with urban analytics, is a more appropriate and all-encompassing neologism. It includes the new techniques, methods, and data hybrids found in current research across quantitative geography, both within and beyond the city.

Regardless that “everything now is ‘new’” suggests a need for academic work to critically assess how this novelty can be managed.

Often, when an area of inquiry is truly “new,” analytical perspectives arise to explain how new practices resolve some core insufficiency of past scientific practice. While large (possibly overambitious) research programs have had their drawbacks (Ramsey, 1926; Zach, 2007), one program has been incredibly influential in statistics over the last 50 years. Tukey’s (1962) “The Future of Data Analysis,” the now-famous rumination on things rotten in the state of mid-century statistics, has been vindicated by the emergence and success of a new “data science” following its principles (Donoho, 2017). This rise of data science in urban analytics and city science is clear as well, but inherits its lack of clarity and sense of newness.

Before this, Geocomputation—a similarly ambitious research program that shifted the momentum in geography (Harris et al., 2017), also called at the time “computational geography”—set out a set of specific conceptual and technical “open problems” upon which researchers could focus their efforts. For instance, Gahegan (2000) defined Geocomputation’s open problems to involve “inclusion of geographical ‘domain knowledge’” within computational tooling or the “development of robust clustering algorithms.” Batty (2019a) suggests that urban analytics might use existing momentum in the domain as a slingshot to find a further new “city science” that “moves beyond data analysis *per se*” (405). Into what, however, is the somewhat vague usual promise of social science: “applications to problems relating to the big questions of our time, inequality, aging, the future of work” (405). The core questions of our new domain are still unclear as the roads forward are still being paved.

To this end, a panel at the 2019 American Association of Geographers meeting convened to interrogate this terminological confusion in urban analytics and set out an agenda for the future of the field. Specifically, the panel aimed to define “Urban Data Science” (UDS) and identify core research questions or research practices for the field. In doing so, the panel explored the delicate interweaving of data science with urban studies and social theory that defines this domain. To guide the panel, we posed a few core questions that help to structure an inquiry into the structure and the future of UDS:

1. Higham et al. (2017) aim to provide a general sense of what is different about work in “city analytics.” When you see work in UDS, what makes it distinct from work in spatial statistics, geocomputation, or urban studies?
2. The recent book by Singleton et al. (2017) offers a data-focused view of urban analytics, suggesting that “Urban Analytics is the practice of using new forms of data in combination with computational approaches to gain insight into urban processes” (15). What new kinds of data or computational approaches are distinctive for UDS?
3. Donoho (2017) provides “six divisions” of activity for data scientists. What might a similar set of divisions look like for UDS, specifically?[1]
4. The field of geocomputation has defined a rather precise set of challenges that animate the domain (Gahegan, 2000). What might these challenges be for UDS?
5. Batty (2019a) provides a goal “mov[ing] beyond data analysis per se.” What might goals or objectives for UDS that “move beyond analysis” look like?
6. A National Science Foundation Advisory Committee (Ramaswami et al., 2018) developed a report on building a “Sustainable Urban Systems Science.” How is the broader ethical, social, or political role of UDS distinct from that which has gone before?

Below is an abridged and synthesized account of the discussion around these questions. In it, we find strong agreement with Batty’s (2019a) desire to “move beyond analysis,” with some suggesting social theory and others suggesting qualitative work as necessary for a successful "urban data science." In other areas, discussants grapple with the notion of "newness" about the field, and suggest that "data science" may be a signaling device with very different valence in academia than in industry. Throughout, panelists grapple with the Hume's classic dilemma: what UDS "is" may be clear enough to urban data scientists, but how UDS "ought" to function in a society of sciences (and how urban data scientists can get there) is not as clear. Below, discussant names are shortened to their initials. Where necessary, additional audience participants names are introduced in full.

Defining UDS

LW: How do we know it's UDS when we see it? What makes UDS distinct from spatial statistics, geocomputation, or urban studies?

GB: There was a legal decision about pornography in the 1960s, where the justice [Potter Stewart] said "I know it when I see it." That's very much the case for UDS. We may be doing regional science or spatial statistics, but we might label our work as UDS for career enhancing reasons or to position it correctly for a certain journal or academic context.

WX: I agree; very often, UDS is a branding or signaling term. For example, when I present myself to industry people, I use the UDS branding to confirm that status. But, when talking with people in the humanities, I might use other labeling. So, we adjust our labeling depending on the audience we're presenting to. But, I do think there's an important aspect to this term: "data science" as a discipline exists primarily in industry, rather than academia.

VM: I will say something I would like to see of UDS. Good UDS takes a holistic approach to science. It might combine spatial statistics, geocomputation, urban studies, but it focuses on the research questions that matter to cities. Further, many of the datasets in UDS are highly private. So, UDS must also take into account privacy concerns and ethical considerations. UDS is the combination of all these methods that makes it distinctive, rather than being a single set of methods. I think this is all necessary to make UDS successful beyond academia. Otherwise, it'll just end up in 10 years being another discipline like the ones we are currently superseding.

SG: I agree. One thing I want to add is that UDS is an emerging domain, and will definitely evolve to incorporate new data streams, latest technologies, methods from spatial statistics or geocomputation, machine learning, and more broadly artificial intelligence. The field itself is also evolving.

AP: I like the term/discussion around "signaling." With so many overlapping terms, is UDS something that we use internally? Or, is it mostly useful for external signaling? I don't use UDS when talking to other academics, but when I talk to an engineer, UDS suddenly makes more sense. For me, the "data" word is uncomfortable. It puts too much emphasis on the data. I don't think UDS should be defined by the *data*; it should be defined by the important foundational research questions as a starting point.

WX: A lot of these new data science degrees have been started because "data scientist" is a popular industrial job title. Schools are starting to recognize that they have to train students to understand common industrial analytical perspectives to meet industrial demand for labor. And, often in industry, a data scientist is not directly driven by a research question. Instead, they're driven by business objectives or they are driven by the data availability (especially in cartographic/locational intelligence applications). Data scientists in industry have a firehose of data. They know that there is something interesting to do with it, but face challenges figuring out how to do that thing.

GB: This is definitely a drawback of existing work in data science in general and UDS in particular. In the past, it's focused on the low-hanging fruit, these "novel" datasets or new kinds of data rather than putting the research question first, and selecting a methodology and a dataset best suited to answer it. So, we often end up not doing those steps of building an intentional dataset. I think most data scientists aren't even trained to develop those datasets, from a community the "old fashioned way" even if it's the best way to answer the question.

What are the goals for UDS?

LW: Batty (2019a) discusses the definition of urban analytics and provides a goal of "moving beyond data analysis *per se*" (405). Batty is discussing the exploration of emergent patterns in urban data, but also the creation of a new science itself. What ambitions do we have for UDS?

VM: I think it is too early to answer. There are a lot of institutions (at least in the US) that are starting Masters and undergrad programs in data science. Maybe in 10 years we'll see if data science has become a stable discipline. Maybe then we'll know a little better what a "data scientist" is, the same way we know what a computer scientist is now, but we did not one hundred years ago.

Regardless, for UDS to be successful broadly, it must be a more holistic package than conventional data science is right now. UDS should combine both qualitative and quantitative methods in a single domain. I'm not a qualitative researcher, but, if we want UDS to succeed and be useful for cities, we must take residents into account. And, not just data about the residents; we need to take into account how the residents feel about technologies that are thrown at them when cities call themselves "smart cities." We need to understand whether residents feel comfortable with those technologies and the data that are being collected from them. If we do not have resident input, UDS cannot succeed. It must be backed by the people themselves. I am not sure if qualitative methods are adjacent to UDS or a part of it at the moment, but I think this is an important part of making UDS successful going forward.

WX: I love that ambition to have qualitative methods. In my own work, I find it very important to integrate qualitative methods. When searching for causal mechanisms, we may need more than data. I really agree with Batty on this point: we should think about what we use data science for. Batty (2019a) suggests we need to move beyond data science toward a "theory of the urban." We need to have theory to drive the questions that we ask and the methods that we apply to our questions. That theory underlies the foundation of UDS, and the theory is not necessarily there yet. He also says we should improve cities, which I agree, but we are not there yet as a domain.

GB: I like the idea of data science plus qualitative data, but qualitative methods do belong adjacent to the domain, rather than inside of it. Data scientists should work with qualitative researchers on mixed methods projects. But, we cannot label everything as "UDS" just because there's empiricism, otherwise the term becomes meaningless. For example, ethnography is clearly not a data science methodology, but it could be very useful for complementing the results of a data science project. Too often, a limitation of UDS is that it puts the science at the forefront and becomes *scientism*, idolizing the methodology rather than accepting that there are multiple ways of knowing about phenomena that can't always be quantified.

AP: Yes, geographers bring this linked perspective to the table. But, in the larger social and policy realm, geographers are not even at the table. When smart cities initiatives are being deployed, it is more likely that computer scientists and engineers are at the table instead. As a discipline, we must push to be in those conversations. In that way, the signaling that "data science" does in UDS can be a powerful way to get us to the table where decisions about these technologies **VM** mentions are made.

SG: I see disagreement about goals being primarily about research paradigms, such as top-down versus bottom-up approaches. We definitely can improve our knowledge of how to integrate distinct research paradigms in planning and development decisions. Especially in community engagement for urban studies, data science might be important. New data visualization methods are becoming much more available, with data tools developed by companies and organizations. This changes the process of decision-making itself, so we do indeed have more to do with public engagement as a discipline.

Is there a limit to UDS?

JAMES SAXON, University of Chicago: Do you want to restrict the ambitions or UDS a little bit? There has to be a point where data science ends and then the analytics begins. For example, if you define data science as only the munging (*the process of changing data into another format*) and the computation, then it becomes like spectroscopy. It's publishable, useful, hard, scientific, and bounded: at some point that science ends. Instead, if UDS is informed by hard scientific questions *and* it also requires you to have a qualitative framework, how does UDS stake a claim? How does it chunk out a place as a distinct "new" academic entity and not a continuation of any other study of the city ever conducted? Is there a cutoff point where you can stop UDS and let some other science start?

WX: Yes, probably, there is. But, there is also something distinct about the tools that are used in data science. I may have the same kinds of questions and may even be trying to use the same methods as spatial statistics (such as geographically weighted regression), but doing that on a *massive scale*, replicating this method in parallel across clusters becomes materially different from what might be considered usual practice in spatial statistics.

AP: To get back to where UDS starts and stops, some of us may see that our responsibilities stop when you have completed the regression analysis and explained the coefficients. But, what about taking that and then help building the theories about urban or social processes? That's a very important part that is difficult to retain once we start chopping up responsibilities very finely and strictly designating which kind of scientist does what.

GB: I do not think that there is a clear stopping point for UDS. The reason for this fuzziness goes back to the impossibility of defining UDS as a label. It has come to mean a constellation of things that we do. But, when I am munging my dataset down into something I can run a regression on, I usually cannot hand it off right at that point. I do some exploratory analysis, I go back and change some feature engineering or operationalization; there is a reciprocal relationship between parts of this larger expansive notion of UDS. I wouldn't say that one part is statistics and another part is data science. I'd say that part of it is data cleaning and part of it is feature processing/engineering, part of it is model building. I don't want to use data science to describe all of it, but I guess we have to.

This is why I do not often describe myself as an urban data scientist unless I am talking to a very general audience. It so often precludes the theory part **AP** mentions. UDS is sometimes the dilettante practice of doing things with software and data about living in cities, the "I can just do something with this dataset I found online." But, for UDS to be a *science*, it both has to *build from theory* and *give back to theory*. Science is a theory-building operation. We often end data science with the patterns or the webmap, but there is more to science than that.

What do urban data scientists do differently from those who have come before?

LW: Donoho (2017), in defining the expansive field of "greater data science," provides six divisions that describe what data scientists do during their research:

1. Data gathering, preparation, and exploration,
2. Data representation and transformation,
3. Computing with data,
4. Data modeling,
5. Data visualization and presentation,
6. Science about data science

Is there anything about these activities that an urban data scientist must change? Is there anything missing from these activities?

VM: UDS is all of the six Donoho suggests, because UDS is data science applied to cities. The activities are the same, but the focus is distinct. The data will reflect some type of behavior or characteristics of the city, the representation, computation, data modeling, visualization, and meta-science will be the same. Some of our defining challenges are also faced when combining large heterogeneous datasets with traditional data sources.

WX: Or, we have so much data available and you know that there's some kind of interesting observation or pattern that you can extract from that available data. We increasingly have more data than we have time to analyze well. Data science becomes a little bit fetishistic because there's so much to *do*. That's why this field has been so driven by the data.

SG: Geographic data science is focused on the domain of geography or geosciences in general. Spatial, and geospatial, are distinct, since it doesn't necessarily need to be on the "Earth." So that's one general distinction. Computer scientists typically call "spatial computing" what we might call geocomputation. But, we don't really want to continue creating artificial distinctions, despite the fact that these characteristics arise from the data itself. Compared with small data scenarios, big data are so large that we can't really load it all into memory at once. This is why [big data] involve a ton of data filtering/sampling issues regarding the 5Vs' characteristics of big data (Anuradha, 2015).

Back to the original question, this is where UDS can really guide research design. We can advise on tool design and implementation, so that users can have guidance on their own large datasets to find results. This is where academic and industrial UDS has made a lot of advances and will continue to.

AP: We have so much data, but we have so little data, too, especially for particular questions. This is a mismatch between the data available and the open questions in our field. As a field, we are now getting past this initial honeymoon phase of being in awe of these initial analytical possibilities. So, now we can remove "new" or "novel" from many of our definitions. I know the Singleton et al. (2017) definition "using new forms of data," has been discussed. Maybe we can remove that "new" at some point.

We also have a responsibility to ensure that the *old* forms do not get defunded. That's a pressure right now with traditional census data in decline. Why would you pay $5m when you can go on Twitter? Of course, we know there is a reason, so it's important to keep this in mind.

WX: I agree; having analyzed/evaluated these new datasets, I can say that 99% of this new data is not usable. It is difficult to ground-truth. What does it mean that you've got this number of observations at this time? You know it's not representative because you know the population of people you bought this data from, but what ways it is unrepresentative are not clear. Ultimately you have to still rely on Census data to relate to something you know is an "unbiased" true population.

GB: Yes, since most of this new data is sensed or automatically harvested, there's no reference dataset to start quantifying biases. A make or break future for UDS will be figuring out what biases are in these datasets, because we are currently forced to take this on faith. This practice reproduces poverty, unequal power dynamics, yet we toss up our hands and say "it's better than nothing!" But, maybe it's not?

WX: Increasingly, I'm skeptical whether it *is* better than nothing. There is no good way to evaluate these new data. How does this private sample change over time? Who is the sample representing over geography and over time? Often, those are not questions that we can answer yet; those are questions that the people who give us the data can not even answer yet.

AP: Further, in many cases, large "new" datasets are no longer free. Maybe they used to be free a decade ago, but now they may be extremely expensive. Access is only for the happy few, and more and more we rely on datasets that are basically controlled by a handful of really large corporations. That's really scary to me. The example of urban movement studies, common in UDS, shows this well: how many of us can afford access fine-grained cell phone data? I think the answer is "very few of us."

VM: But, some cities are already ahead of us there. Some cities are requiring bike share companies to put a certain number of bikes in certain areas for equity concerns or ensure that companies provide an API for researchers that is open and free (or, at least publishing origin–destination data). If there are UDS projects deployed in a city, I think that city and its politicians and citizens ought to have a say about what type of data needs to be shared by companies with researchers in an open and re-usable way. Because these applications are specifically urban, then city governments do have a say.

Moving beyond big data

VM: Further, the panel is focusing a lot on the large scale that these types of datasets provide, but we don't necessarily need large scale data to do good science. Electoral scientists have long predicted election outcomes accurately using a small number of people in surveys. In contrast, many in UDS are idealizing the information we can extract from large datasets. If we have a small dataset, but it is representative of the population at large, maybe we can avoid using large unrepresentative datasets with no ground-truth. Maybe the solution is to use new technologies to make smaller but higher quality datasets with fewer ethical complications. Maybe the results will be even better if we do this instead of focusing on massive data byproducts.

AP: Maybe this is one aspect where urban research is different. Questions on the scale of the nation may indeed be answerable by a sample of 1000 people. But, consider neighborhoods in most cities. If we want to make sure that we have enough observations in each small neighborhood, then all of a sudden, sourcing data becomes a different question than a nationally representative survey.

SG: And more and more studies are now trying to combine designed research datasets, volunteered, and passively collected data, because the passive data are not necessarily an intentional output. It is usually a byproduct that can be transformed into use for a scientific application. So, this mixture helps us derive a ground-truth for other "non-designed" datasets.

GB: If we interpret UDS to be the broader pipeline from data collection all the way down to activism, then there are a ton of new things we can do. Policymakers are wringing their hands about interventions in gentrification or displacement crises. If you speak with housing planners, they often don't know what rents are like in their city. They're waiting, for example, for five-year American Community Survey (an ongoing survey by the U.S. Census Bureau) data to come out and give them massively lagged information about these problems. If there's some way for us to collect information about that for housing planners, then maybe they can intervene in crises proactively. That's kind of a different take on data science, but if we accept this broader pipeline, then this becomes an important role for us: fixing problems where they actually exist.

SG: Yes. One task in data science is predictive modeling/knowledge discovery. I view this as a well-grounded aspect of UDS from traditional urban science theory. But, maybe for data science, it's more "exploratory process modelling." Empirically, we find "something good" and then review theory to see if our results explain something we already know. If we find something new that we do not already know, then that is progress. It can be challenging to see that from the outset of a study, but this kind of exploration can be exciting.

WX: Another distinction between different types of models in data science is the predictive model versus generative model. Can you derive inferences from the data? Can you extract the underlying parameters that are influencing your outcomes? This is important for UDS, in the sense that we don't want to *just* predict, we want to understand the relationships between different factors in the city. This will be what drives the underlying policy.

SG: I'd also like to point out that the cities and urban data portals are providing a new platform for cities to build on top of. They're making aggregated public datasets publically available. They also push along UDS as an academic discipline because we get free and clean access to that data for our research and can push back on companies or provide insights to government. I'm not sure if they'll use it or not, but I think this is a cycle that can push openness in government at a local level.

VM: Relating to what we can and cannot (or should and should not do), something Donoho (2017) says about metascience is: "The true effectiveness of a tool is related to the probability of deployment times the probability of effective results when deployed" (756).

The point I was making before about qualitative data science boils down to: we are doing good UDS if we can achieve Donoho's (2017) points 1–5 (which are not necessarily new for many domains), but also if we succeed at doing science about UDS. We need to make sure that whatever systems we build are adopted by decision-makers in the city, which is the whole point of calling it UDS.

WX: But there is a distinction between "just" data science and "urban" data science. The biggest distinction is the spatial component. When you have spatial data, that changes a lot of the activities. How do you create spatial indices over terabytes of data or visualize that on a map? There are unique challenges here that have to be addressed by urban data scientists.

AP: What Donoho (2017) is missing for us is theory! To call back to where data science starts and stops, some of us may see that our responsibilities stop when you've completed the regression analysis and explained the coefficients. But, what about taking that and then help building the theories about urban or social processes? That's a very important part that is difficult to retain once we start chopping up responsibilities.

GB: Data science must be embedded in a discipline, so "urban" data science research must be embedded within the broader conversations in urban studies. Often, the practice of data science is freelance, though. However, urban data scientists ought to have read the canon, read some critical literature, and they ought to be able to speak to stakeholders. UDS is often a professionalized field, which also poses its own sorts of problems. So if we can't speak to the open challenges in city planning practices or urban geography—like **AP** was saying, the human and ethical implications of social research—I'm not sure that UDS is safe to do because lives are at stake.

We do not necessarily need more quantitative information about the impact of freeways on urban space or urban renewal on human lives, we need political solutions. Too often, the low-hanging fruit for us is to clean up the data we have and run some regressions so that we might be able to help some policymaker someday maybe not tear a freeway through that impoverished neighborhood. But too often, we're ignoring the human angle, even if we're thinking about the ethical perspectives of our work. We're not often working in the community when we're doing data science. We're not often asking the data science questions we *should* ask on behalf of the people whose lives might be impacted by the answer.

Conclusion and future directions

However hard it will be for urban analytics to "mov[e] beyond data analysis" to "city science" (Batty, 2019a), it will no doubt be harder for a "fetishistic" UDS to do so. Panelists consistently suggested that a successful UDS is impossible without a broader and more successful *urban studies*. The thirst for theory is clear and manifests in the panel's many charges for a broader normative debate in UDS. What UDS ought be according to this set of early career scholars, it seems, is expansive, integrated, and interdisciplinary. Where UDS may be now, at least in its worst forms, is the "dilettante practice" of "naive and reductionist" (Kitchin, 2013: 266) science on "accidental" data (Arribas-Bel, 2014). Where the panel (and this larger discussion) fails is in explaining what makes for a best form of UDS.

As such, this panel report aims to initiate further normative thinking about what quality UDS (and indeed, city science) might look like. Is all good UDS integrated with qualitative research? The panel remains divided whether this belongs "within" or "alongside" the domain, and a boundary remains uncertain. What does current UDS do differently from existing work? This is not clear. Instead, all panelists agreed on the broader ethical requirements incumbent upon good UDS, but did not volunteer criteria by which these requirements are judged or a strategy by which an urban data scientist might satisfy them. Is the domain even "new?" No, and some on the panel believed asking that question may not even be productive. What is clear is that there are clear "winners" and "losers" in the problems of the past (such as those noted by Gahegan (2000)); the idea that "many of the struggles associated with scaling algorithms up are transitory" (Donoho, 2017: 764) means that computational feasibility (at least in some areas of UDS) is now no longer a headlining concern.[2] What the headlining research concerns are beyond individual research agendas are still not clear.

Thus, if anything, this panel demonstrates the fundamental instability in terms that Batty (2019b) provides: there is a subtle yet important miscommunication about what, exactly, is being studied. Only in closing remarks (removed for brevity here) do panelists begin to engage specifically with how "new" UDS changes the fundamental practice of their science specifically, let alone how it changes urban studies generally. While the panel makes the same "greater" and "lesser" pivot as Donoho (2017) to suggest that the full potential of city science is not yet realized, it is unlikely that UDS could yet be identified conclusively. The panel suggests that UDS is a superset of past work on the city, but the contents of this superset "depends on what the meaning of the word 'is' is." While negative archetypes in the field were easy for the panel to offer, a positive vision is still occluded. It seems likely that good social science remains conceptually similar to past work—Mertonian middle-range theory-driven empirical study of socially important phenomena—but results in the development of "projects that would not merely have been impossible otherwise, but which we would have likely not even imagined" (King, 2014). Good UDS will require better urban theory, but the rest still remains up for debate.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. One attempt to answer this question contemporaneous to our panel is found in Singleton and Arribas-Bel (2019). We regret the inability to include it in the panel guidance.
2. Although, in the complete discussion, **WX** notes that this is true *if and only if* your analysis is simple enough, or if it is widely used and engineers of large-scale computing projects are pressured to implement it. The ongoing corporate consolidation of high-performance computing and machine learning frameworks may yet bring Gahegan's (2000) concern about "disabling technologies" back to the fore, as high-performance computing becomes optimized only for the specific analytical paths of concern in industrial applications.

References

Anuradha J (2015) A brief introduction on big data 5Vs characteristics and hadoop technology. *Procedia Computer Science* 48: 319–324.

Arribas-Bel D (2014) Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography* 49: 45–53.

Batty M (2019a) Urban analytics defined. *Environment and Planning B: Urban Analytics and City Science* 46(3): 403–405.

Batty M (2019b) On the confusion of terminologies. *Environment and Planning B: Urban Analytics and City Science* 46(6): 997–999.

Beynon MJ, Crawley A and Munday M (2016) Measuring and understanding the differences between urban and rural areas. *Environment and Planning B: Planning and Design* 43(6): 1136–1154.

Boeing G (2019) Urban spatial order: Street network orientation, configuration, and entropy. *Applied Network Science* 4(1): 67.

Borruso G (2003) Network density and the delimitation of urban areas. *Transactions in GIS* 7(2): 177–191.

Breuer J, Walravens N and Ballon P (2014) Beyond defining the smart city. Meeting top-down and bottom-up approaches in the middle. *TeMA – Journal of Land Use, Mobility and Environment* 7. DOI: 10.6092/1970-9870/2475.

Chodrow PS (2017) Structure and information in spatial segregation. *Proceedings of the National Academy of Sciences* 44: 11591–11596.

Deavers K (1992) What is rural? *Policy Studies Journal* 20(2): 184–189.

Donoho D (2017) 50 years of data science. *Journal of Computational and Graphical Statistics* 26(4): 745–766.

Evans KL, Chamberlain DE, Hatchwell BJ, et al. (2011) What makes an urban bird? *Global Change Biology* 17(1): 32–44.

Frias-Martinez V, Soto V, Hohwald H, et al. (2012) Characterizing urban landscapes using Geolocated Tweets. In: *2012 international conference on privacy, security, risk and trust and 2012* *international conference on social computing*, pp.239–248. Washington, DC. IEEE Computer Society.

Gahegan M (2000) What is geocomputation? A history and outline. Available at: http://www.geocomputation.org/what.html (accessed 2 March 2019).

Gao S, Yu L, Yaoli W, et al. (2013) Discovering Spatial Interaction Communities from Mobile Phone Data. *Transactions in GIS* 17(3): 463–481.

Hall SA, Kaufman JS and Ricketts TC (2006) Defining urban and rural areas in U.S. epidemiologic studies. *Journal of Urban Health* 83(2): 162–175.

Harris R, O'Sullivan D, Gahegan M, et al. (2017) More bark than bytes? Reflections on 21+ years of geocomputation. *Environment and Planning B: Urban Analytics and City Science* 44(4): 598–617.

Higham D, Batty M, Bettencourt LMA, et al. (2017) An overview of city analytics. *Royal Society Open Science* 4(2): 161063.

Hoggart K (1988) Not a definition of rural. *Area* 20(1): 35–40.

Isserman AM (2005) In the national interest: Defining rural and urban correctly in research and public policy. *International Regional Science Review* 28(4): 465–499.

King G (2014) Restructuring the social sciences: Reflections from Harvard's institute for quantitative social science. *PS: Political Science & Politics* 47(1): 165–172.

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

Kitchin R (2016) The ethics of smart cities and urban science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374(2083): 20160115.

Krupat E and Guild W (1980) Defining the city: The use of objective and subjective measures for community description. *Journal of Social Issues* 36(3): 9–28.

Louail T, Lenormand M, Cantu Ros OG, et al. (2014) From mobile phone data to the spatial structure of cities. *Scientific Reports* 4: 5276.

McDade TW and Adair LS (2001) Defining the 'urban' in urbanization and health: A factor analysis approach. *Social Science & Medicine* 53(1): 55–70.

Naik N, Raskar R and Hidalgo CA (2016) Cities are physical too: Using computer vision to measure the quality and impact of urban appearance. *American Economic Review* 106(5): 128–132.

Nam T and Pardo TA (2011) Conceptualizing smart city with dimensions of technology, people, and institutions. In: *Proceedings of the 12th annual international digital government research conference: Digital government innovation in challenging times*, Dg.o '11, pp.282–291. New York: ACM.

O'Brien DT, Sampson RJ and 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.

Pryor RJ (1968) Defining the rural-urban fringe. *Social Forces* 47(2): 202–215.

Ramaswami A, Bettencourt L, Clarens A, et al. (2018) *Sustainable urban systems report*. National Science Foundation, Sustainable Urban Systems Committee. Available at: https://www.nsf.gov/ere/ereweb/ac-ere/sustainable-urban-systems.pdf (accessed 28 January 2019).

Ramsey FP (1926) Mathematical logic. *The Mathematical Gazette* 13(184): 185–194.

Shelton T and Poorthuis A (2019) The nature of neighborhoods: Using big data to rethink the geographies of Atlanta's neighborhood planning unit system. *Annals of the American Association of Geographers* 109(5): 1341–1361.

Singleton A and Arribas-Bel D (2019) Geographic data science. *Geographical Analysis in Analysis*. Epub ahead of print. DOI: 10.1111/gean.12194 (accessed 4 April 2019)..

Singleton A, Spielman S and Folch D (2017) *Urban Analytics*. Thousand Oaks, CA: SAGE.

Small C, Pozzi F and Elvidge CD (2005) Spatial analysis of global urban extent from DMSP-OLS night lights. *Remote Sensing of Environment* 96(3): 277–291.

Tannier C, Thomas I, Vuidel G, et al. (2011) A fractal approach to identifying urban boundaries. *Geographical Analysis* 43(2): 211–227.

Taylor L (2011) No boundaries: Exurbia and the study of contemporary urban dispersion. *GeoJournal* 76(4): 323–339.

Thompson EM (2016) What makes a city 'smart'? *International Journal of Architectural Computing* 14(4): 358–371.

Tsai Y-H (2005) Quantifying urban form: Compactness versus 'sprawl.' *Urban Studies* 42(1): 141–161.

Tukey J (1962) The future of data analysis. *The Annals of Mathematical Statistics* 33(1): 1–67.

Williams S, Xu W, Tan SB, et al. (2019) Ghost cities of China: Identifying urban vacancy through social media data. *Cities* 94: 275–285.

Wilson JQ (1968) The urban unease: Community vs. city. *The Public Interest* 12: 25–39.

Yin J, Soliman A, Yin D, et al. (2017) Depicting urban boundaries from a mobility network of spatial interactions: A case study of Great Britain with geo-located twitter data. *International Journal of Geographical Information Science* 31(7): 1293–1313.

Zach R (2007) Hilbert's program then and now. In: Jacquette D (ed.) *Philosophy of Logic*. *Handbook of the Philosophy of Science*. Amsterdam: North-Holland, pp.411–447.

Wei Kang is a Postdoctoral Scholar at the Center for Geospatial Sciences at the University of California Riverside.

Taylor Oshan is an Assistant Professor in the Center for Geospatial Information Science at the University of Maryland, College Park.

Levi J Wolf is a Senior Lecturer (Assistant Professor) in Quantitative Human Geography at the University of Bristol & Fellow at the Alan Turing Institute.

Geoff Boeing is an Assistant Professor in the Department of Urban Planning and Spatial Analysis at the Sol Price School of Public Policy, University of Southern California.

Vanessa Frias-Martinez is an Assistant Professor at the University of Maryland Institute for Advanced Computer Studies, and an affiliate assistant professor in the Department of Computer Science.

Song Gao is an Assistant Professor of Geographic Information Science at the University of Wisconsin-Madison.

Ate Poorthuis is an Assistant Professor at the Singapore University of Technology and Design.

Wenfei Xu is a PhD student at the University of Columbia and former Data Scientist at Carto.com.