September 2019

SNS RESEARCH BRIEF *nr*58

Cities and Big Data

NEW "BIG DATA" sources allow for the measurement of city characteristics and outcome variables at higher collection frequencies and more granular geographic scales than ever before. However, big data will not solve large urban social science questions on its own. Big urban data has the most value for the study of cities when it allows for the measurement of the previously opaque or when it can be coupled with events that are unforeseen by people or places.

THE LACK OF CONNECTION between the urban applications of social science and the physical city has partly been driven by a lack of data on the physical attributes of urban spaces. Big data has the potential to bridge this gap.

IT IS SHOWN HOW GOOGLE STREET VIEW IMAGES can be used to predict income in New York City, suggesting that similar imagery data can be used to map wealth and poverty in previously unmeasured areas of the developing world. The images can also be used to estimate tenants' valuation of, e.g., housing characteristics.

CROWDSOURCED DATA from online platforms such as Yelp are often contemporaneous and geographically finer than official government statistics. Evidence is presented showing that Yelp data can complement government surveys by measuring economic activity such as opening new restaurants in close to real-time at a granular level and on almost any geographic scale.

AUTHOR

Edward Glaeser
Fred and Eleanor Glimp Professor of Economics, Harvard University

SNS RESEARCH BRIEF | The objective of the SNS Research Brief series is to make research accessible for decision-makers in politics, business and the public sector. Financial support has been granted from the Jan Wallander and Tom Hedelius Foundation. Responsibility for the analysis and the conclusions in the research briefs rests with the authors alone.

Design: Patrik Sundström / Allan Seppa

Introduction

Bridging the gaps between urban economics and the physical city.

Historically, there has been a divide between urban economics and the physical aspects of the city itself. Social scientific research has been detached from subjects such as architecture and streetscapes. This lack of connection has been driven in part by a lack of data on the physical attributes of urban spaces. The "big data" revolution will change this. Big data turns a cross section of space into living data, offering a broader and finer picture of urban life than has ever been available before. Moreover, in combination with predictive algorithms, big data may allow us to extrapolate outcome variables such as house prices or income to previously unmeasured parts of a population. This policy brief showcases some examples of how big data can be used to develop cities.

To answer classic social science questions such as understanding the impact and mechanisms of urban growth and how the physical city interacts with social outcomes, big data becomes powerful once it is combined with exogenous sources of variation. Exogenous means something that is out of the control of the decision maker, such as bad weather or a new policy. In urban contexts, the two key exogenous variation sources are "shocks to places" and "shocks to people"; the former consists of high frequency events that affect geographic regions (e.g., the opening of large manufacturing plants; see Greenstone, Hornbeck, and Moretti 2010), while the latter consists of high-frequency events that affect geographic regions within cities (e.g., the moving to opportunity [MTO] experiment; see Chetty, Hendren, and Katz 2016; Katz, Kling, and Liebman 2001). Information on exogenous events causes these observational settings to mimic the classic statistical laboratory situation where a researcher assigns a certain "treatment" to an object being studied. This alleviates issues otherwise common to observational studies,[1] enabling the separation of causality from correlation. Causal knowledge is useful for, e.g., the design of policies.

1. Often called endogeneity issues, e.g., reversed causality, omitted variable bias and self-selection.

The strongest case for using big data to assess economic outcomes in cities is when finely detailed geography can be matched with longitudinal data[2] and exogenous events that are tied to a particular locale. In those settings, big data can enable researchers to examine whether there is a causal effect on people in the proximity of the exogenous event, no matter where they move.

Big data is also improving city management. By making their operations more data driven, cities can fine-tune regulations, improve the allocation of scarce resources, and forecast future needs. Crucially, for many urban data interventions, simply being able to predict outcomes or characteristics is valuable on its own even without understanding the underlying phenomena causing the outcomes. Moreover, many data-driven interventions are scalable; hence, the expansion of data collection and digitization efforts across cities attracts entrepreneurship and innovation.

Overall, the findings discussed in this policy brief are relevant in the following ways. Policymakers benefit from having real-time snapshots of how neighborhoods and the local economy are evolving over time, which is useful for both planning and policy purposes. Potential investors and homebuyers also benefit from understanding, for example, how housing prices are likely to change in different neighborhoods over time. In addition, real estate investors might want to know where housing prices are likely to increase. The scale and magnitude of urbanization make the questions discussed here relevant on a global scale.

Together with my coauthors Hyunjin Kim, Scott Duke Kominers, Michael Luca and Nikhil Naik, I have studied these topics in Glaeser et al. (2017), Glaeser, Kim, and Luca (2017) and Glaeser, Kim and Luca (2018). The present policy brief summarizes these findings in the aforementioned studies. These studies use U.S. data, but many takeaways herein are valid in the European Union (EU) context as well.

2. Longitudinal data collects repeated measurements on the same subject, e.g., individuals' incomes for two years or more.

Quantification and City Policy

In addition to their value for urban science, improvements in the quantification of cities can dramatically change the way that cities operate and evaluate policies.

New ways in which big data sources are impacting city policies are becoming abundant: governments are digitizing and sharing records, and private firms are collecting high-frequency measurements of local businesses, traffic, and other urban features. In this section, a taxonomy of the new types of urban data now available to researchers and policymakers is presented, followed by a discussion on how the new data can impact policy.

Taxonomy of Data Sources

Digital Exhaust

One valuable but underutilized source of data is *digital exhaust*, which is the trail of data left online through everyone's day-to-day use of the Internet. Across a variety of domains, digital exhaust can help measure the physical city. Review platforms such as Yelp and TripAdvisor provide direct measures of the quality of services and establishments throughout cities worldwide. Social media platforms such as Twitter and Facebook can inform us about the pulse of neighborhoods or about the structure of social networks. LinkedIn can shed new light on labor markets and search costs.[3] Search queries from platforms such as Google and Bing contain insight about the needs and preferences of a physical city. Zillow provides new insight into housing markets, as does data from sharing platforms such as Airbnb.

Digital exhaust data can be directly applied to city management. Yelp reviews, for example, provide detailed, high-frequency data on restaurants that can be used to assess hygiene (see the extended example discussed below, as well as Glaeser et al. 2016; Kang et al. 2013), and Google searches can be used to help predict flu outbreaks (see, e.g., Carneiro and Mylonakis 2009; Ginsberg et al. 2009; Polgreen et al. 2008; Yang, Santillana, and Kou 2015).

3. For an overview of these and other user-generated content platforms, see Luca (2016).

Open Government Records

Thirty years ago, cities stored most of their records on paper. Now, a growing digitization movement seeks to convert data that were historically on paper to electronic, machine-readable records. Digitized records are often made available online. For example, criminal records have been publicly available for several decades now in many U.S. states yet were very difficult for policymakers and researchers—much less the public—to access. Over time, criminal records have slowly become digitized and readily available, which has made research easier, but it has also influenced incentives for recidivism and criminal behavior more generally (Finlay 2009; Luca 2016).

Your digital exhaust can be used to improve the city.

An "open data" movement seeks to increase transparency by making cities' internal data publicly available. Many large cities (e.g., Boston, Chicago, Copenhagen, San Francisco and Stockholm) have created open, freely accessible "data portals" that researchers and citizens can use to access digitized records; many of these data portals are updated in real time. The availability of open data encourages entrepreneurs to look for ways in which city data can be used to enhance welfare and creates possibilities for new partnerships between city officials and researchers.

Corporate Data

Private data from companies represent a third but less developed approach to measuring the physical city. In addition to the digital data mentioned above, one can envision using gym memberships to understand health behavior, College Board data to gain additional insight into student performance, and credit card transactions to quantify changes in spending over time. Furthermore, data from cell phone carriers such as Vodafone and Telia could shed light on the patterns by which people move around the city.

How Can New Data Empower the City?

At a basic level, cities specialize in three activities that are deeply reliant on—or can be greatly improved through—data and analysis.

1. Cities evaluate and enact policies and regulations.
2. Cities operate public services.
3. Cities forecast future activities for the sake of planning and policymaking.

Next, I briefly describe ways in which big data can influence each of these activities.

Policy Evaluation

Historically, most empirical policy analyses have looked at relatively narrow outcomes. One might examine, for example, the impact of hotel tax rates on the prices of hotels. It would be more beneficial to measure the broader impact of taxes, but historically, other factors such as the impact of taxes on "quality" would be very difficult to measure. Now, tax changes, TripAdvisor ratings, Priceline prices, and Airbnb listings can be combined to obtain a much broader view of the intended and unintended consequences of tax policy on the physical city.

In addition to broadening the outcomes under consideration, new data can lead to higher frequency estimates of changes. Suppose, for example, that someone wants to evaluate the impact of unemployment benefits on job searches. Traditional analyses might examine the length of unemployment and the average income after one year. However, LinkedIn and similar sites could in principle give us measures of day-by-day job search behavior.

Operating Public Services

While research has traditionally focused on policy evaluations, there is a growing acknowledgment of the practical importance of predicting problems. Cities are responsible for allocating scarce resources. For example, cities choose which domestic violence cases to follow up on and which labor market complaints to investigate; in both settings, the underlying choice problem here is not a program or an evaluation but rather a prediction problem. The city must predict which domestic violence offenders are likely to reoffend and which labor market complaint is most likely to unearth a serious issue. Using data to improve these predictions about the physical city creates value—and new data sources are central to this task (see Kleinberg et al. 2015).

User ratings can be used to allocate scarce resources.

Another example concerns the cell phone carrier Telia, which helped the Helsinki Regional Transport Authority identify hotspots to deploy "feeder" bus lines to the subway in the Helsinki/Espoo area. This made it more convenient for commuters to choose public transportation, resulting in a car traffic decrease of 8 percent during the period of November 2017 to January 2018.

Forecasting

Urban planners and policymakers forecast future economic activity through time-series analyses on leading indicators of activity. New data sources—especially coupled with machine learning—have the potential to revolutionize forecasting. Zillow, TripAdvisor, and LinkedIn, for example, provide measurements that can be used in estimating future housing prices, tourism, and unemployment. Data from app-based payment systems can provide insight into upcoming retail spending and consumption patterns.

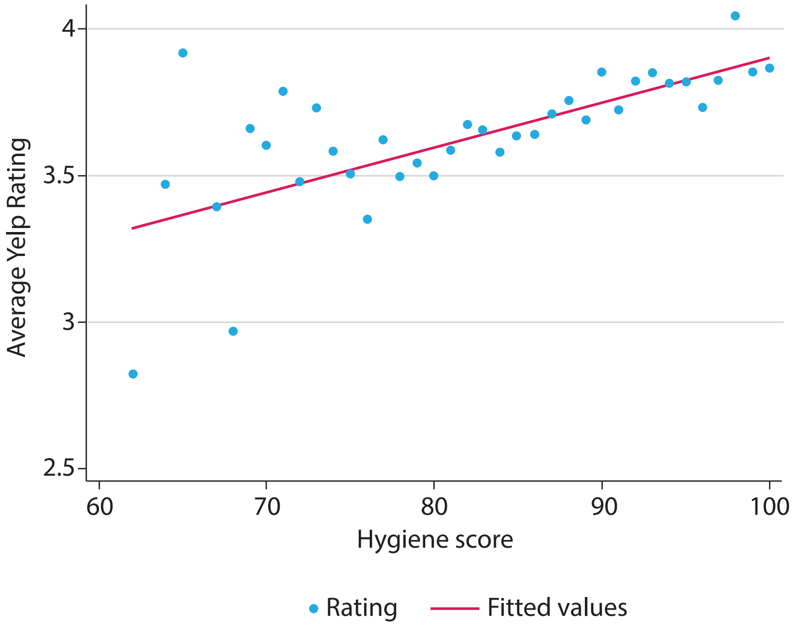
Closing Example: Data-Driven Hygiene Inspections

This section concludes with an applied example showing how data and predictive tools can directly inform city resource allocation.

In nearly every developed country, health inspectors examine restaurants to identify unsafe restaurant practices (such as storing food at unsafe temperatures), which can lead to foodborne illness. These inspectors are typically allocated according to the perceived health risk posed by a restaurant or cuisine. For example, a sushi restaurant may be inspected more often than a burger joint because sushi is more likely to lead to food sickness. In addition, health inspector allocation is effectively random.

However, inspections do not *have* to be random. Suppose instead that we were to base the likelihood of inspection on evidence from Yelp reviews. Perhaps we would start with a search on Yelp for terms such as "sick" or "dirty"; we would probably find a few culprits, but a predic-

Figure 1. Correlation between Yelp Ratings and Hygiene Inspection Services.



Notes: Review data consists of reviews on Yelp.com for restaurants in San Fransisco, CA from September, 2010 through September , 2013. Hygiene scores are from the San Fransisco Department of Public Health, for the same timeframe.

tive algorithm trained through machine learning can do much more than that. The algorithm would "learn" from the histories of reviews and inspection outcomes and then predict the likelihood of finding violations based on more recent reviews. Inspectors could then be reallocated to restaurants that are most likely to have violations.

Kang et al. (2013) and Glaeser et al. (2016) explored the feasibility of using natural language processing to predict hygiene violations with reviews on Yelp. To see how powerful even a simple adjustment can be, consider Figure 1, which shows the correlation between Yelp ratings and hygiene scores. Even before applying machine learning techniques, it is clear that Yelp scores can help identify hygiene scores. In a competition run with the city of Boston to develop a predictive algorithm for Boston, Glaeser et al. (2016) demonstrated that using Yelp reviews to guide inspections may significantly increase the number of health risks identified without increasing costs.

However, as discussed in the introduction, urban policy questions relying on a concrete understanding of a causal effect must combine the data with information on exogenous shocks. Other policies, however, can be informed directly through data, often in combination with predictive algorithms.

Measuring the Streetscape

In this section, it is demonstrated how applying computer vision algorithms to Google Street View data (or other image corpora) can be used (1) to measure the physical characteristics of neighborhoods and (2) to estimate neighborhood income.

In the past decade, Google Street View has extensively photographed the built environment in more than 100 countries. Almost all American cities have been documented in high definition, and the resulting images can be classified using computer vision algorithms. Some of these images can be linked to corresponding GPS-coded attributes of interest (such as housing prices, income, or ratings of urban upkeep). This information could create a training data set that could be utilized to train algorithms.

Images can be used as data…

... to predict incomes and housing prices.

After training, the algorithms are shown images of houses not seen before (called a test sample) to judge whether the training went well. If it did, then the algorithm is likely to make good predictions of house prices, incomes in the area and so on from pictures of houses in new settings, e.g., other cities. The potential of this procedure is shown below. Two examples of the procedure will be given in these two distinct but related questions:

1. Can Street View imagery data be used to predict income?
2. Can Street View data improve predictions of how certain housing characteristics affect housing prices?[4]

The first question is most relevant for developing countries, where we sometimes have large image corpora but no reliable large-sample income data. The second question is likely to be more relevant in the developed world, where we have price data but typically have not used visual images of the streetscape as explanatory variables. Predicting prices with streetscape imagery may also have public policy value as a tool in property value appraisal in places (both developed and developing) where governments rely on property taxes.

If images are available for an entire city, a computer vision model trained on a small sample of income data can produce a citywide map of wealth and poverty, as well as measures of income segregation. If we have images at different points in time, we can then test how individual interventions change the distribution of wealth.

4. E.g., can it improve the quality and fit of hedonic regressions? Hedonic regression is a technique that can be used to determine how different housing characteristics affect housing prices. Here, the price of, e.g., an apartment, is assumed to depend on several characteristics such as the size of the apartment, whether it is a penthouse or not, the distance to the closest commuting station and in which part of the city the apartment is located. The hedonic regression approach can be used to estimate, e.g., the market valuation of such characteristics, i.e., how the characteristics influence the market price of the apartment.

Predicting Income and Housing Prices with Pixels

As a proof of concept, it is demonstrated that the median income of residents in New York City can be predicted from Street View images using a computer vision model. Further, it is shown that the computer vision model trained to predict median incomes from images of New York City is able to predict the median income of images from Boston with almost the same accuracy as in New York. Finally, predicted income is linked with housing prices to show the potential use of this technology in hedonic housing price regressions. The procedure is described in more detail in Glaeser et al. (2017).

The main results from Glaeser et al. (2017) can be summarized as follows:

- The fit of the model works well in contexts other than the training and test settings.
- The income measure predicted from images alone captures 77 percent of the variation in the true income measure.

A model trained on New York City images works well for Boston. However, keep in mind that Boston and New York are reasonably similar places. Further, the analysis here is done using block group data rather than individual address data.

Hedonic Pricing

We now turn to our next exercise: linking images to prices. In this case, the interest lies in the extent to which physical attributes can add predictive power to models when the aim is to predict housing prices. In some cases, the interest may just be in expanding predictive power—perhaps if the government is trying to improve an automated appraisal process for property tax purposes. In other cases, we may actually want to know which physical attributes explain differences in housing values. Street View imagery may be useful in both settings. For example, computer vision technology has enabled the identification of particular street-level attributes, such as potholes. In principle, these attributes can be added to a hedonic price model. At present, however, the focus is on the simpler task of just predicting prices with pixels.

In essence, we are currently aiming to answer the question of whether the physical attributes of neighborhoods that attract rich people also increase housing prices. The main results can be summarized as follows:

- The predicted income from the images has significant explanatory power for housing prices, both in the training sample and test samples.
- The things that can be seen from the street (i.e., in the pictures) have approximately as strong a predictive power for the housing prices as the things that cannot be seen from the street.
- Similar results hold true for predicting housing prices in Boston using the New York City trained model.

Thus, we conclude that Google Street View can predict income in New York (and Boston using the New York-trained model), and predicted income helps us predict housing prices in our sample. This does not mean that we can predict income well in the developing world, but it does provide some hope that Google Street View and similar products will enable us to better understand patterns of wealth and poverty worldwide.

Nowcasting Gentrification

Gentrification can be loosely defined as a process where deteriorated urban neighborhoods are renovated due to the inflow of relatively wealthy residents. This may result in a number of consequences, such as the displacement of restaurants and convenience stores, implying a rearrangement in the distribution of workplaces and often a change in ethnic composition. Hence, it is highly relevant for policy to be able to measure gentrification. This section suggests a way of doing so based on street imagery.

Data Description

Our first measure of gentrification is via the housing price data provided by the Federal Housing Finance Agency (FHFA). These data represent an annual repeat sales index for over 18,000 five-digit ZIP codes in the United States, described in Bogin, Doerner, and Larson (forthcoming). We use data from 2012 to 2016, and the average real growth of this index over this period is 3.1 percentage points.

Three measures of neighborhood demographics are used. These are available for five-year windows from the American Community Survey (ACS): percent college educated, percent aged between 25 and 34, and the share of the population that is white. Because education tends to be reliably correlated with both income and housing costs, the percent of people with a college education in an area provides a reasonable metric to measure gentrification. In our sample, the average ZIP code in New York showed that the number of adults with college degrees increased by 2.6 percent.

The measurement of neighborhood change informs policymakers.

Our final measure of neighborhood change is the change in StreetScore, drawn from Naik et al. (2017). This measure contains information on how respondents rated images from Google Street View on perceived safety. These ratings were used as training data for computer vision techniques, which generated StreetScores for more neighborhoods. The StreetScore measure is interpreted as a proxy for the overall physical quality of the neighborhood rather than safety per se.

For measures of changes in business categories, we use data from Yelp's business listings that are sourced through user submissions, business owner reports, partner acquisitions, and internal data quality checks.[5]

Results on Local Housing Prices

We first explore the ability of Yelp data to predict contemporaneous changes in housing price growth at the ZIP code level, looking at the period from 2012 to 2016 by following Rascoff and Humphries (2015), who link the proximity of a Starbucks to the price growth on Zillow. In our

5. Despite the granularity and availability, Yelp data have limitations that are discussed in further detail in Glaeser, Kim, and Luca (2017). Yelp's business classification is assigned through user and business owner reports, which results in unsystematic industry categorization that does not correspond to government datasets. Furthermore, the quality of Yelp data depends on the degree of Yelp adoption, which has grown over time. Given these issues, we only count businesses as open if they have received at least one recommended Yelp review.

version, we examine whether price growth is correlated with contemporaneous growth in the number of Starbucks cafes, which allows us to understand whether the addition of a Starbucks is an indicator of gentrification.

First, the correlation between the growth in home prices and the increase in the number of Starbucks in the same ZIP code during the same year is calculated. A one-unit increase in the number of Starbucks in a given year is associated with roughly a 0.5 percent increase in housing prices. This effect is large, but the explanatory power of the Starbucks measure is modest. Further, the direction of causality in this relationship is a priori unclear. For example, Starbucks may target the opening of its cafes in places that are on the upswing, so the correlation may reflect Starbucks' strategy, that is, higher incomes in a ZIP code area may cause Starbucks to open establishments there. On the other hand, businesses may contribute to gentrification; here, a newly opened Starbucks may lead people with higher incomes to move into the ZIP code area, implying that the causality would go in the opposite direction.

To partially distinguish between these hypotheses, we control for unobserved differences between ZIP codes that do not change over time. Our time period is short, so if Starbucks is targeting growing areas, then this approach should eliminate much of the correlation. Here, the 0.5 increase falls to 0.17. Another analysis includes both the current and past measures of Starbucks growth, as well as the change in the number of Starbucks that are closed and the increase in the number of Starbucks reviews. The increase in the number of Starbucks reviews is also predictive of neighborhood change: a ten-unit increase in the number of reviews is associated with a 1.4 percent increase in housing prices in the ZIP code. Since the presence of a Starbucks is less important than whether the community reviews the Starbucks, this finding challenges the interpretation that people are paying for the proximity to a Starbucks. While Starbucks may be a particularly prominent coffee shop, it is not the only possible retail establishment that may signal gentrification at the local level. Therefore, the analysis is expanded to include all of the cafes listed in Yelp over the same time period. This change increases the number of ZIP codes because more ZIP codes have at least one cafe during this time period. Similar results are found, although the magnitude of the power of the cafe reviews is somewhat weaker than for Starbucks. The difference between Starbucks and the cafe results lends some support to the upscaling hypothesis.

In Glaeser, Kim, and Luca (2017), we expand our analysis to other industries that Yelp has classified. In many cases, similar to Starbucks, the number of Yelp reviews provided additional predictive power beyond just the addition of a business, suggesting that both changes in the local economy and changes in the use of Yelp are related to gentrification.

Results on Demographic Change

We now explore whether the local business ecosystem shifts with demographic changes in a neighborhood. The focus is on New York City, and we examine demographic changes between 2007–2011 and 2012–2016, as well as StreetScore changes between 2007–2014.

The change in the number of grocery stores is statistically significantly correlated with the change in the number of adults with college degrees. The correlation of the change in the number of grocery stores with the age and racial composition of the ZIP code is also statistically significant, but this correlation is approximately one half the size of the correlation with the change in the percent that are college educated. These results seem compatible with the literature on "food deserts" that refers to how poorer people live in areas with fewer healthy food options.

The number of laundromats is correlated with the share of the population that is young, which is perhaps less surprising. As laundromats are rarely "upscale," this result seems more compatible with business densification.

Correlations exist between the change in the share of the population that is college educated and changes in the number of cafes, bars, restaurants, barbers, wine bars, convenience stores, fast food restaurants, florists, and restaurants categorized by Yelp as being pricey. Restaurants, barbers, and florists also correlate with the number of people who are young. Correlations with the racial composition were almost uniformly weaker.

In Glaeser, Kim, and Luca (2017), we reproduce these results for Boston, Chicago, Los Angeles, and San Francisco and examine correlations with the number of Yelp reviews by category. Many of the patterns are broadly similar, with two significant differences. The number of laundromats no longer has a strong correlation with gentrification. In the other four cities, unlike New York, several of the Yelp review counts correlate strongly with the number of younger people in the ZIP code, potentially due to the geographic variation in the age of Yelp reviewers.

The final measure is of the physical change in the neighborhood as measured by StreetScore. As before, we begin with New York City and then turn to other large urban areas in Glaeser, Kim, and Luca (2017). To keep the results comparable, we continue to look at the ZIP code level data, although there is no reason why we could not look at the block itself. At the ZIP code level, the strongest correlation is with the number of vegetarian restaurants, which had a much weaker correlation with the change in the share of the college educated. The second strongest correlation is with the change in the number of Starbucks, and the third strongest is with wine bars. This mirrors our previous results with regards to demographic change.

Concluding discussion

Big data is particularly valuable when it can improve policymaking directly. The ways in which Yelp (and other crowdsourced rating tools), e.g., can augment city services such as sanitary inspections have been discussed. More generally, big data offers the ability to use the broader civil society to augment the functions of government by lowering the costs of contributing to government services. Smart phone apps and other similar accessories provide tools that citizens can use to give feedback to governments quickly and inexpensively. In this research, the potential for measuring gentrification and predicting citizens' income as well as housing prices in cities was showcased.

In general, big data can help cities in meaningful ways and improve research on cities but only if it is used with thoughtful care. Big data will do far more for urban research if it is paired with exogenous sources of variation, and it will do far more for policymaking and implementation if it is paired with openness to new methods.

Big data can improve policymaking directly.

Ongoing discussions about how to update or replace the national census across many countries have been witnessed in recent years. For example, the United Kingdom considered replacing the census with administrative data as well as third-party data from search engines such as Google (Hope 2010, Sanghani 2013). One of the areas that the U.S. Census Bureau has been considering in its new plan to pare $5.2 billion dollars from its cost of $20 billion for the decennial census is to utilize administrative records and third-party data (U.S. Census Bureau 2015, Mervis 2017).

Third-party data can augment official statistics.

Our analyses of a possible data source, Yelp, suggest that these new data sources can be useful complements to official government data. Yelp can help predict contemporaneous changes in the local economy. It can also provide a snapshot of economic change at the local level. Yelp is a useful addition to the data tools that local policy-makers can access.

Ultimately, data from platforms such as Yelp – combined with official government statistics – can provide valuable complementary datasets that will ultimately allow for more timely and granular forecasts and policy analyses, with a wider set of variables and a more complete view of the local economy.

An important question in relation to this work is how the trade-off between privacy for individuals and the openness of the rich information contained in data sources such as those discussed here should be handled. This is especially relevant in light of the General Data Protection Regulation (GDPR) act implemented in the EU during 2018.

Additionally, the question of how public statistical agencies should handle and/or complement data generated by private firms is of great importance.

References

Bogin, A.N., W.M. Doerner, and W.D. Larson. Forthcoming. "Local House Price Paths: Accelerations, Declines, and Recoveries." *Journal of Real Estate Finance and Economics*.

Carneiro, H. A., and E. Mylonakis. 2009. "Google Trends: A Web- Based Tool for Real-Time Surveillance of Disease Outbreaks." *Clinical Infectious Diseases*, 49(10), 1557–64.

Chetty, R., N. Hendren, and L. F. Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *American Economic Review*, 106(4), 855–902.

Finlay, K. 2009. "Effect of Employer Access to Criminal History Data on the Labor Market Outcomes of Ex-Offenders and Non-Offenders," in *Studies of Labor Market Intermediation*, edited by D. H. Autor. Chicago: University of Chicago Press, 89–125.

Ginsberg, J., M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant. 2009. "Detecting Influenza Epidemics Using Search Engine Query Data." *Nature*, 457(7232), 1012–14.

Glaeser, E.L., S. D. Kominers, M. Luca, and N. Naik. 2017. "Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life." *Economic Inquiry*, 56(1), 118-37.

Glaeser, E. L., A. Hillis, S. D. Kominers, and M. Luca. 2016. "Crowdsourcing City Government: Using Tournaments to Improve Inspection Accuracy." *American Economic Review,* 106(5), 114–18.

Glaeser, E.L., H. Kim, and M. Luca. 2017. "Nowcasting the Local Economy: Using Yelp Data to Measure Economic Activity." National Bureau of Economic Research Working Paper 24010.

Glaeser, E.L., H. Kim, and M. Luca. 2018. "Nowcasting Gentrification: Using Yelp Data to Quantify Neighborhood Change." *AEA Papers and Proceedings*, 108, 77–82.

Greenstone, M., R. Hornbeck, and E. Moretti. 2010. "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings." *Journal of Political Economy*, 118(3), 536–98.

Hope, C. 2010. "National Census to be Axed after 200 Years." *The Telegraph*, July 9, 2010. http://www.telegraph.co.uk/news/politics/7882774/National-census-to-be-axed-after-200-years.html. Accessed July 6, 2017.

Kang, J. S., P. Kuznetsova, M. Luca, and Y. Choi. 2013. "Where *Not* to Eat? Improving Public Policy by Predicting Hygiene Inspections Using Online Reviews," in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA: ACL, 1443–48.

Katz, L. F., J. R. Kling, and J. B. Liebman. 2001. "Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment." *Quarterly Journal of Economics*, 116, 607–51.

Kleinberg, J., J. Ludwig, S. Mullainathan, and Z. Obermeyer. 2015. "Prediction Policy Problems." *American Economic Review*, 105(5), 2015, 491–95.

Luca, M. 2016. "User-Generated Content and Social Media," in *Handbook of Media Economics*, edited by S. Anderson, J. Waldfogel, and D. Strömberg. Amsterdam, The Netherlands: North Holland.

Mervis, J. 2017. "Scientists Fear Pending Attack on Federal Statistics Collection." *Science Magazine*, January 3, 2017. http://www.sciencemag.org/news/2017/01/scientists-fear-pending-attack-federal-statistics-collection. Accessed July 6, 2017.

Naik, N., S.D. Kominers, R. Raskar, E.L. Glaeser, and C.A. Hidalgo. 2017. "Computer Vision Uncovers Predictors of Physical Urban Change." *Proceedings of the National Academy of Sciences* 114 (29), 7571–76.

Polgreen, P. M., Y. Chen, D. M. Pennock, F. D. Nelson, and R. A. Weinstein. 2008.

"Using Internet Searches for Influenza Surveillance." *Clinical Infectious Diseases*, 47(11), 1443–48.

Rascoff, S. and S. Humphries. 2015. "Confirmed: Starbucks Knows the Next Hot Neighborhood before Everybody Else Does." *Quartz*, January 28, 2015. https://qz.com/334269/what-starbucks-has-done-toamerican-home-values/. Accessed January 4, 2018.

Sanghani, R. 2013. "Google Could Replace National Census." *The Telegraph*, June 26, 2013. http://www.telegraph.co.uk/technology/google/10142641/Google-could-replace-national-census.html. Accessed July 6, 2017.

U.S. Census Bureau. 2015. "2020 Census Operational Plan Overview and Operational Areas." https://censusproject.files.wordpress.com/2015/12/2020-census-opplan-conference-call_the-census-project_10-21-15_final-1.pdf. Accessed July 6, 2017.

Yang, S., M. Santillana, and S. C. Kou. 2015. "Accurate Estimation of Influenza Epidemics Using Google Search Data via ARGO." *Proceedings of the National Academy of Sciences of the United States of America*, 112(47), 14473–78.