

## How Should Urban Planners Be Trained to Handle Big Data?

Steven P. French, PhD., FAICP  
Dean and John Portman Chair  
College of Architecture  
Georgia Institute of Technology  
Atlanta, GA 30332-0155  
[steve.french@coa.gatech.edu](mailto:steve.french@coa.gatech.edu)

Camille Barchers  
Graduate Student  
School of City and Regional Planning  
Georgia Institute of Technology  
Atlanta, GA 30332-0155  
[cbarchers3@gatech.edu](mailto:cbarchers3@gatech.edu)

Wenwen Zhang  
Graduate Student  
School of City and Regional Planning  
Georgia Institute of Technology  
Atlanta, GA 30332-0155  
[wunwunchang@gmail.com](mailto:wunwunchang@gmail.com)

July 15, 2014

# How Should Urban Planners Be Trained to Handle Big Data?

Steven P. French, Camille Barchers, and Wenwen Zhang

School of City and Regional Planning

Georgia Institute of Technology

Atlanta, GA 30332-0155

Keywords: Urban Planning, Analytics, Education, Visualization, Simulation

Abstract: Historically urban planners have been educated and trained to work in a data poor environment. Urban planning students take courses in statistics, survey research and projection and estimation that are designed to fill in the gaps in a data poor environment. For decades they have learned how to use the census data, which is comprehensive on several basic variables, but is only done once per decade so is almost always out of date. More detailed characteristics are based on a sample and are only available in aggregated form for larger geographic areas.

But new data sources, including distributed sensors, infrastructure monitoring, remote sensing, social media and cell phone tracking records, can provide much more detailed, individual, real time data at disaggregated levels that can be used at a variety of scales. We have entered a data rich environment, where we can have data on systems and behaviors across smaller time increments and with a greater number of observations on a greater number of factors (Few, 2009; Lohr, 2012). Planners are still

being trained in methods that are suitable for a data poor environment (Contant & Forkenbrock, 1986; Cuzzocrea, Song, & Davis, 2011; Kaufman & Simons, 1995). In this paper we suggest that visualization, modeling, data mining and machine learning are the appropriate tools to use in this new environment and we discuss how planning education can adapt to this new data rich landscape. We will discuss how these methods can be integrated into the planning curriculum as well as planning practice.

Planning methods have been the source of much discussion over the past few decades.

Practitioners and researchers have examined what methods planning schools teach and how these are used in practice. The suite of traditional methods courses taught in planning programs – inferential statistics, economic cost-benefit analysis, sampling, and research design for policy evaluation-- remains largely stagnant, despite a rapidly changing reality in which planners are expected to work. Although the focus of this paper is on the impact of big data for planning methods, other variables have also contributed to the need for additional methods to tackle planning problems. The rise of ubiquitous computing and a hyper-connected communication network as well as new private investment in data collection have created an environment in which greater amounts of data exist than ever before. The ability of the planner to analyze and use this data is no longer limited by computing power or the cost of data collection, but by the knowledge that planners possess on data analytics and visualization techniques.

Educating planners with skills that are useful for practice has been a key tenant of many planning programs over the years. Several studies have been conducted to understand better how planning programs succeed or not at this goal. Surprisingly, the most recent comprehensive investigation of planning education and skills demanded by practitioners was conducted in 1986. In this survey, four important conclusions are identified as relevant to how planners were being educated and the professional skills they would be required to use (Contant & Forkenbrock, 1986). They found that the methods taught in planning programs remained highly relevant to the methods needed for practicing planners, and the authors concluded based on their survey results that planning educators were adequately preparing their students to solve planning problems in practice. They cited communication skills (writing and speaking) and analysis and research

design as critical components of planning education and practice, but noted that educators needed to remain vigilant on seeking relevance (Contant & Forkenbrock, 1986). The piece also identified several changes occurring throughout the 1980s that affected the planning profession—the rise of micro-computing and the expansion of methods being offered by planning schools. Contant and Forkenbrock (1986) wrote “...there is little to suggest that planning schools are overemphasizing analytic methods, nor do they appear to be failing to any real extent in meeting the demands of practitioners interviewed. While more techniques are required than these practitioners feel that all planners should understand, it certainly is arguable that this situation is not at all bad.” That survey of methods is now almost thirty years old, and new realities exist that require educators to expand the methods taught in planning schools (Goodspeed, 2012; Sawicki & Craig, 1996). Yet, despite wide acknowledgement of the changing data landscape, planning curricula still resemble their traditional form. Kaufman and Simons completed a follow-up to this investigation which surveyed planning programs specifically on methods and research design. The more limited focus on this 1995 study “revealed a rather surprising lack of responsiveness among planning programs over time to practitioner demand for [quantitative research methods]” and that “planning programs do not seem to teach what practitioners practice, and not even what practitioners should practice” (Kaufman & Simons, 1995). In a 2002 study focused on the use of technology within planning programs, Urey claims that the haphazard approach with which planning programs have introduced the use of technology to serve larger goals (research, analysis, modeling) might be problematic as increased microcomputing power becomes more widespread. While manual techniques serve learning objectives within planning methods courses, the use of technology is now required (Urey, 2002). This leaves planning educators today with two questions relevant to big data and methods: what

new methods must we now include in our curriculum, and what technology must students become familiar with to employ these methods in an ethical, accurate, and precise way? Given these questions, we reviewed current methods requirements at planning schools to assess whether or not planning programs have begun to respond to these questions and adapt to the changing data landscape.

In a non-scientific review of methods taught at the top ten planning schools (as listed by Planetizen [<http://www.planetizen.com/education/planning>]), we discovered that almost all programs require that planners be trained in statistics, economic cost-benefit analysis, and research design. Of the programs reviewed, including MIT, Cornell, Rutgers, UC Berkeley, University of Illinois Urbana Champaign, UNC Chapel Hill, University of Southern California, Georgia Institute of Technology, UCLA, and University of Pennsylvania, none required students to seek additional data analysis courses outside of the department. Although the review of these programs was not scientific and limited to information published online for prospective students, it does generally persuade us that planning education has yet to see value in teaching planners methods largely reserved for those in the fields of computer science and engineering. We argue, as Contant and Forkenbrok argued thirty years ago, that relevance of planning programs to planning practice is important. Contant and Forkenbrok reminded educators to be vigilant in their understanding of skills that are in demand for practitioners— yet we have failed to do this in regards to our methods curricula.

Big data, although currently a popular topic, is not new—and the concept of big data dates back to 2001, when industry analyst Doug Laney articulated the definition of big data as any data set

that was characterized by the three Vs: Volume, Velocity and Variety (Laney, 2001). Big data sets are characterized by containing a large number of observations, streaming and fast speed and requiring real time analytics. Big data sets are also usually mixed format, joined by a common denominator, but not all of the same type. In sum, any data sets that are too large and complex to process using conventional data processing applications can be defined as big data.

Several pioneers in the industry have already started to process and analyze big data (Lohr, 2012). For instance, UPS now tracks 16.3 million packages per day for 8.8 million customers, with an average of 39.5 million tracking requests from customers per day. The company stores more than 16 petabytes of data. Through analyzing those datasets, UPS is able to identify real time on-road traffic conditions, daily package distribution patterns and together with the latest real time GIS mapping technology, the company is able to optimize the daily routes for freight. With all the information from big data, UPS has already achieved savings in 2011 of more than 8.4 million gallons of fuel by cutting 85 million miles off of daily routes (Davenport & Dyché, 2013). IBM teamed up with researchers from the health care field to use big data to predict outbreaks of dengue fever and malaria (Schneider, 2013). It seems that big data, together with advanced analysis and visualization tools, can help people from a wide variety of industries explore large, complex data sets and extract patterns that were once very difficult to unveil. Given the increasing use of big data across fields that share interests with the field of city planning, planners should more deliberately explore and develop methods for using big data to develop insights about cities, transportation patterns and the basics of urban metabolism.

Data analytics, as a powerful tool to investigate big data, is becoming an interdisciplinary field. There are new programs at universities across the United States that aim to teach students how to

grapple with big data and analyze it using various tools. For this paper, we collected and reviewed some common tools and skills that are taught in data analytics courses. We gathered course information from John Hopkins, Massachusetts Institute of Technology, University of Washington, and Georgia Institute of Technology. We noted that machine learning/data mining and data visualization are the tools that are frequently taught in the programs to prepare students to handle big data and some of them are actually quite new to urban planners.

Machine learning is a core subarea of artificial intelligence. Machine learning uses computer algorithms to create explanatory models. There are different types of learning approaches, including supervised learning, unsupervised learning, and reinforcement learning. Although some of the terminologies may be completely new to planners, the actual methods turn out to be quite familiar. For example, the regression model is one of the methods that is frequently used in supervised learning process. Planners who work with remote sensing images often apply supervised classification methods to reclassify the images into land cover images based on various bands of the image. However, planners may not be familiar with other machine learning methodologies or algorithms, such as unsupervised learning and reinforcement learning.

Unsupervised learning tries to identify regularities (or clusters or groupings) in the input datasets without correct output values provided by the supervisors. Reinforcement learning is primarily used in applications where the output of the system is a sequences of actions (e.g. playing chess). In this case, what's important is not a single action, but a sequence of actions that will achieve the ultimate goal. If the machine learning methods are applied to large databases such as big data, then it is called data mining. Data mining tries to identify and construct a simple model



with high predictive accuracy, based on the large volume of data. The model is then applied to predict future values.

Most of the programs we reviewed also include data visualization components to help data analysts better communicate their final data products. Some data visualization techniques, such as multivariate data representations, table and graph designs are quite conventional. However, those techniques may also be applied in innovative ways to help convey information behind data in a clearer manner. One example is the information graphics or infographics, which improve human cognition by utilizing graphics to improve the visual system's ability to extract patterns and trends (Few, 2009; Smiciklas, 2012). The latest trend in data visualization is to take the advantage of webs to present data in an interactive way. To effectively present big data interactively, the designer needs to be equipped with knowledge regarding how human beings interact with computers, and how different interaction types (i.e. filtering, zooming, linking, and brushing) will affect human being's cognition ability.

In addition to the core courses, these new interdisciplinary programs require the students to master at least one program or query language. SQL is a popular requisite and, in a survey on tools for data scientists, over 71% of respondents used SQL (King, and Magoulas, 2013). Some programs also require students to understand and use open software such as R and R studio.

While these methods for analyzing data may seem completely out of place within a planning methods framework, they actively seek to create ways in which researchers can describe, explore, and explain data. These categories of data analysis are described in depth in Earl Babbie's *Survey Research Methods*. This text serves as one of many fundamental introductions to methods for planners, and by grouping the new suite of tools available to planners and data scientists within these categories, planners can see how these tools might be useful to them. For

example, data visualization is one of the key ways in which data scientists are exploring big data sets (Few, 2009). Data visualization acknowledges that our typical methods of data exploration (descriptive statistics, graphing, and the like) are ill equipped to handle larger data sets, and even less equipped to communicate information from those data sets to the public and to decision makers. By introducing planners to the growing field of data visualization, we can expand their ability to not only to use larger data sets but to communicate the information garnered from those data sets. As the basis for research, exploration of data sets will allow planners to ask additional questions. These additional questions will require explanatory analysis, and within this group of methods, tools such as machine learning and data mining will help planners generate predictive models from larger data sets.

Many data sets that planners have to deal with in the future may turn out to be big data. Credit card data or web browsing histories may help planners to predict the focus of emerging public concerns. As a matter of fact in MIT's big data courses, there is a case study regarding how to utilize the Google search records to estimate the trends within the real estate industry (MIT, 2014). Social media, such as Twitter and Facebook, have already become powerful information sources regarding almost every aspect of social life. Analysis of twitter feeds can help to identify the extent and intensity of hazard events. There are already studies towards how to utilize information extracted from Facebook's friend list to forecast the use of airplanes. GPS or real time transportation information can help planners to calibrate and develop more accurate activity based travel demand models to forecast future travel patterns. Moreover, the real time information about energy flows such as water, sewer, and electricity flows may equip planners with critical information to design more energy efficient and sustainable cities to make built

environment more resilient to natural hazards and climate change. Planning is defined by our special affinity for place based issues, and this focus on place will be one of the critical ways in which typical data sets can become “big data.” Location is the ultimate relational field, and our ability to link data sets through location will create big data sets that are especially useful to planners. If location is the ultimate relational tool, than planning data sets will only continue to increase in size, speed, and format in the future. The importance of teaching planners how to accurately and precisely examine and explore this data cannot be understated, yet, our work to prepare this paper leads us to believe that planning programs have not yet taken the steps required to introduce these methods to planning students.

The big data analysis tools, such as machine learning and data visualization, can help planners to make better use of the big data sets. The Memphis Police Department used machine learning and data mining approaches to predict potential crime based on past crime big data, the serious crime rate was reduced by approximately 30%. The city of Portland, Oregon optimized their traffic signals based on big traffic data, and was able to reduce more than 157,000 metric tons of CO<sub>2</sub> emissions in six years (Hinssen, 2012). In sum, the machine learning techniques can help planners to analyze the future development of urban areas in a more accurate way to eliminate or at least ease some the current urban problems. The explanatory power of machine learning will be critical for planners seeking to use big data to solve long-term challenges in cities and communities.

Data visualization has always been considered useful in the planning process, primarily as a communication method. However, it is now critical in exploring large data sets. Data visualization can help planners better understand how people live, work and behave within urban

context. When paired with more explanatory tools such as machine learning, data visualization becomes a critical step in the planning process. Visualization can also continue to be used as a way for planners to convey their planning concepts to corresponding stakeholders during the public participation process. In this way, visualization is used as an interpretation toolkit to help people digest the complex analysis results from big data. Planners continue to be more comfortable using traditional graphs, tables, and animation images to visualize their results. However, some planners are now using more advanced web based tools to display the information in interactive ways to encourage public participation. This trend has been on the rise for some time, and the demand for practitioners with visualization skills continues to increase (Few, 2009; Goodspeed, 2012; Sawicki & Craig, 1996).

We argue in this paper that planners would benefit greatly from the introduction of more advanced methods of descriptive, exploratory, and explanatory data analysis in order to more effectively use an ever increasing amount of relevant data. However, some key questions for additional research emerged as we prepared this paper. Two characterizations of planning prove to be challenging when introducing new methods to the planning student and practicing planner—the field of planning is inherently place based, and as such it has the potential to take typical planning data and transform it into big data by introducing volume and mixed format information into databases. Secondly, the field of planning is concerned with the long-term. Thus far, big data has been analyzed to provide insights into short term challenges—the speed that is required to transform a data set into big data means that we are frequently examining real time data at such minute discrete time fragments that it verges on becoming continuous. As planners, we need to be asking a larger question that relates to not just what methods can be used to employ this data, but how this data can be employed in our search for long-term solutions. How

does the use of minute by minute Twitter text analysis related to planning issues allow us to reframe planning issues for years to come? How does second to second transportation data give us a new way to understand transportation patterns that matter beyond the day and into the next decade? Although we did not set out to answer these questions in this paper, we do believe that posing them along with giving planners insight into how we can use data that is becoming increasingly large, fast moving, and varied, will help us redefine planning methods for the next generation of planning students and practitioners. Big data is an exciting new asset for planners who have always struggled to explore and explain patterns and trends based on limited observations of discrete data—we should make the best use of this data by giving our planners the tools with which to analyze it and communicate it. Like others who have written on the topic of big data in cities, we do caution that data should not be used or analyzed purely for data's sake, that is, planners are tasked with a more nuanced goal than computer scientists: they must find ways in which to use the data to make existing communities better and to provide better solutions than were previously available (Mattern, 2013; Sawicki & Craig, 1996) . In order to help planners achieve these goals, we must revamp our methods offerings in our planning programs to better represent the ways in which we are interacting with big data as researchers.

## References

- Contant, C. K., & Forkenbrock, D. J. (1986). Planning methods: An analysis of supply and demand. *Journal of Planning Education and Research*, 6(1), 10-21.
- Cuzzocrea, A., Song, I.-Y., & Davis, K. C. (2011). *Analytics over large-scale multidimensional data: the big data revolution!* Paper presented at the Proceedings of the ACM 14th international workshop on Data Warehousing and OLAP.
- Davenport, T. H., & Dyché, J. (2013). Big Data in Big Companies. *International Insti.*
- Few, S. (2009). *Now you see it: simple visualization techniques for quantitative analysis*: Analytics Press.
- Goodspeed, R. (2012). The Democratization of Big Data. Retrieved from <http://www.planetizen.com/node/54832>
- Hinssen, P. (2012). *Open Data, Power, Smart Cities How big data turns every city into a data capital* Retrieved from <http://datascienceseries.com/assets/blog/Greenplum-Open-Data-Power-Smart-Cities-web.pdf>
- Kaufman, S., & Simons, R. (1995). Quantitative and research methods in planning: are schools teaching what practitioners practice? *Journal of Planning Education and Research*, 15(1), 17-33.
- Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note*, 6.
- Lohr, S. (2012, February 11, 2012). The Age of Big Data, Opinion. *The New York Times*. Retrieved from <http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html?pagewanted=all>
- Mattern, S. (2013). Methodolatry and the Art of Measure. 2014, from <http://places.designobserver.com/feature/methodolatry-in-urban-data-science/38174/>
- Sawicki, D. S., & Craig, W. J. (1996). The democratization of data: Bridging the gap for community groups. *Journal of the American Planning Association*, 62(4), 512-523.
- Schneider, S. (2013). 3 Examples of Big Data Making a Big Impact on Healthcare This Week. Retrieved from <http://blog.gopivotal.com/pivotal/p-o-v/3-examples-of-big-data-making-a-big-impact-on-healthcare-this-week>
- Smiciklas, M. (2012). The power of infographics. *Using pictures to communicate and connect with your audience. Indiana: Que.*
- Urey, G. (2002). A Critical Look at the Use of Computing Technologies in Planning Education The Case of the Spreadsheet in Introductory Methods. *Journal of Planning Education and Research*, 21(4), 406-418.

