APA
American Planning Association
Making Great Communities Happen

PAS REPORT 585

BIG DATA AND PLANNING

Kevin C. Desouza and Kendra L. Smith

APA RESEARCH MISSION

APA conducts applied, policy-relevant research that advances the state of the art in planning practice. APA's National Centers for Planning—the Green Communities Center, the Hazards Planning Center, and the Planning and Community Health Center—guide and advance a research directive that addresses important societal issues. APA's research, education, and advocacy programs help planners create communities of lasting value by developing and disseminating information, tools, and applications for built and natural environments.

Since 1949, APA's Planning Advisory Service has provided planners with expert research, authoritative information on best practices and innovative solutions, and practical tools to help them manage on-the-job challenges. PAS publications include PAS Reports, PAS Memo, PAS QuickNotes, and PAS Essential Info Packets. Learn more at planning.org/pas.

James M. Drinan, JD, Chief Executive Officer; David Rouse, FAICP, Director of Research; Camille Fink, PHD, Senior Editor.

PAS Reports are produced in the Research Department of APA.

For missing and damaged print issues, contact Customer Service, American Planning Association, 205 N. Michigan Ave., Suite 1200, Chicago, IL 60601 (312-431-9100 or customerservice@planning.org) within 90 days of the publication date. Include the name of the publication, year, volume, and issue number or month, and your name, mailing address, and membership number, if applicable.

ISBN: 978-1-61190-188-7
APA's publications office is at 205 N. Michigan Ave., Suite 1200, Chicago, IL 60601–5927.
APA's headquarters office is at 1030 15th St., NW, Suite 750 West, Washington, DC 20005–1503.
E-mail: pasreports@planning.org

ABOUT THE AUTHORS

Kevin C. Desouza, PhD, is a Foundation Professor in the School of Public Affairs at Arizona State University (ASU). For four years, he served as associate dean for research for the College of Public Service and Community Solutions. He is also a Nonresident Senior Fellow at the Brookings Institution. Immediately prior to joining ASU, he directed the Metropolitan Institute in the College of Architecture and Urban Studies and served as an associate professor at the School of Public and International Affairs at Virginia Tech. From 2005 to 2011, he was on the faculty of the University of Washington (UW) Information School and held adjunct appointments in the College of Engineering and in the Daniel J. Evans School of Public Policy and Governance. At UW, he co-founded and directed the Institute for Innovation in Information Management (I3M); founded the Institute for National Security Education and Research, an interdisciplinary, university-wide initiative in August 2006 and served as its director until February 2008; and was an affiliate faculty member of the Center for American Politics and Public Policy. More information is available at www.kevindesouza.net.

Kendra L. Smith, PhD, serves as a policy analyst at the Morrison Institute for Public Policy at Arizona State University (ASU). Previously, she served as a post-doctoral scholar in the College of Public Service & Community Solutions and research fellow in the Center of Urban Innovation at ASU. She is interested in a wide range of public sector issues, such as innovation, citizen engagement, urban planning, and technology adoption, as well as how the public and private sectors can develop mutually beneficial partnerships to enhance innovation, efficiency, and effectiveness for society. She understands the need for utility and real-world applications for public sector research and seeks to place solutions directly in the hands of public administrators through publications, consultation, and advisement. Her work has been featured in *Stanford Social Innovation Review*, *Governing*, and *Public Management (PM)* as well as on NPR and through the Brookings Institution. More information is available at www.kendrasmithphd.com.

ON THE COVER

Photo from Thinkstock (Vladimir_Timofeev)

TABLE OF CONTENTS

EXECUTIVE SUMMARY

We have more data than ever before on all kinds of things planners care about, ranging from fixed objects, such as buildings, to dynamic activities, such as driving behavior, and at all levels of granularity, from cities to individuals and everything in between. In addition, we have unprecedented access to data in real time, which gives us new capacities to monitor and evaluate the state of objects (artifacts) and agents. We have technology to thank for a great deal of this. Much of what we interact with is fitted with technology sensors, and our own (consumer) technologies—such as mobile phones, tablets, and fitness trackers—are also valuable sensors and emitters of data. Technology is not only generating these data but is also providing us with the computational tools to analyze, visualize, interpret, and act on large amounts of data. In addition, we have tools that are automating decision making through the embedding of intelligence, which is derived through analyses of large amounts of data.

PLANNING IN THE ERA OF BIG DATA

Digitization of all aspects of the city—from sensors embedded in our infrastructure to activity and movement tracking capabilities through GPS and even social sensing (crowdsourced data from sensors closely attached to humans, such as mobile phones or wearables)—allows us to learn about the sentiments of residents regarding policy and social options and provides a distinct opportunity to rethink how we plan and manage cities and communities. Key to this is the ability to leverage data through analytics in an effective and efficient manner.

Real-time analytics applied to data are also increasing our situational awareness about our environments. This increased real-time situational awareness will enable individuals, communities, and organizations to make more intelligent, or smarter, decisions. And, with smart technologies, there are greater opportunities to become aware of early warning signals through real-time access to sensors and predictive models that alert us to problems. The use of data, analytics, and smart technologies will help cities manage their complexities and become less reactive to challenges and threats.

Planners use various types of data and data sources in their complex plan-making activities. Big data has the strong potential to enrich various stages of plan making, including visioning, problem assessment, scenario planning, and plan implementation. Big data empowers planners and decision makers by helping them to better understand current situations and predict future ones more precisely and accurately. The intense amounts of time and labor required for data collection as part of comprehensive planning might be resolved in the near future with big data generated using volunteer sources or created through more formal processes. However, planners will need to be prepared to deal not only with data quality issues but also the institutional challenges of collecting, analyzing, and incorporating big data into planning processes.

WHAT ARE BIG DATA AND ANALYTICS?

A set of data is big when it is too large and too complex to be stored, transferred, shared, curated, queried, and analyzed by traditional processing applications. There is no specific size that is assigned to big data, as it is always growing. A more complex and accurate answer requires the understanding that data can be big in several dimensions. The simplest dimension is volume. We are talking about data that ranges in size from terabytes (1000^4) to petabytes (1000^5), exabytes (1000^6), and zettabytes (1000^7). With advances in computation, what is considered big today will be small in the near future, and what is gigantic today will be medium-sized in the short term.

The velocity at which data are generated and collected has undergone major changes over the last few years. Today we generate data at an accelerated rate across a multitude of environments. The lag between when an activity is conducted and when it is recorded has, for most things, almost disappeared.

The variety of data streams, formats, and types of data that we have access to has also exploded. Today, we have to contend with data that come across multiple streams—from formal systems to informal platforms (e.g., social media)—and formats including text and audio, visual and aural, and even olfactory. These data streams emit data with high degrees of variability in structure, frequency, predictability, and other characteristics.

Analytics is key to harnessing the power of big data. It is the process of connecting, analyzing, transforming, modeling, and visualizing data to discover valuable and actionable information. Two key activities take place during analytics. First, data are summarized and integrated to reduce their volume and converted into higher-level metrics and indicators that enable individuals to process the data. Second, the information generated is situated within the appropriate contexts in order to make sense of the data and conclusions to be rendered.

THE INTERNET OF THINGS

Through smart data and Internet of Things (IoT) technologies, more growth will happen in areas involving variables that previously were difficult to measure. IoT is the interconnection of devices that "talk'" to each other, including personal electronics, sensors, and networks. Wearables, or wearable technology, is an IoT technology made up of clothing and accessories with embedded sensors and software that can connect to other objects or devices without human intervention. These devices will monitor specific conditions, such as motion, vibration, and temperature. The promise of IoT technology is a revolutionized, fully connected world where the environment and people are connected by objects.

For urban planners, IoT opens up a whole new realm of possibilities. Using this information, city infrastructures can be reimagined, which will have impacts on social, political, and environmental policies. Connected cars will make driving safer and cut down on carbon emissions while public spaces can adapt and adjust to users' needs to create entertainment, educational opportunities, and interactive spaces for gathering, collaboration, fellowship, and further innovation. This will all be done through insights gleaned from the collection of big data and analytics, insights that will be used to connect all facets of daily life. This might seem far-fetched now, but imagine 20 years ago, when floppy disks were used to save computer work and people used cameras with actual film that had to be developed in order to share physically printed photos with other people.

ANALYTICAL APPROACHES

Critical to the ability to leverage big data are analytical techniques. Traditional analytical techniques have to scale to handle the volume of data involved when dealing with big data. In addition, analytical techniques must be able to integrate different types of data and operate across a wide assortment of systems and platforms.

Open-Source Tools

The cost of analytics can range from free to expensive. For those planners not looking to break the bank on a new set of computational tools, there are open-source tools that are available at no cost. Open-source tools have source code that is openly published for use or modification from the original design—free of charge.

Data Mining

Data mining helps us discover latent patterns and associations between variables in large datasets. There are a multitude of analytical techniques that can be applied to uncover information from databases, including association rules, decision trees, and classification models. From a process point of view, generic data mining works the following way: the dataset is split into two (the training set and the validation set). Models that are deemed good have good predictive capacity and also have explanatory power (i.e., how much variance in the dependent variable is explained by the variables in the model). Association rules (or association mining) is a data mining technique that is probably the most well-known and straightforward. With this technique, one can discover the presence of an element in a dataset as it relates to the co-occurrence of other elements.

Machine Learning

Analytical approaches use algorithms to iteratively learn from data and build (and update) models. Machine learning approaches normally take one of two learning approaches: supervised or unsupervised. In supervised learning, the model is developed by analyzing patterns using a set of inputs and outputs. The analytical tool continues to learn and the prediction accuracy of the model improves over time. In unsupervised learning, data elements are not labeled and we allow the algorithm to identify patterns among the various elements. The machine normally will have to identify the latent structure in the dataset through a clustering approach to identify groups and clusters. Just as with data mining, several analytical techniques can be used for both supervised and unsupervised learning depending on the size, structure, and format of the dataset.

Sentiment Analysis

Today, given the popularity of social media and the proliferation of commentary on these platforms, sentiment analysis (or opinion mining) has become a popular analytical approach. At its core, sentiment analysis leverages techniques in natural language processing, computational linguistics, and text analysis to uncover information from large bodies of text. Machine learning approaches that focus on sentiment analysis are often used when assessing polarity. Opinions—gathered from sources such as tweets and Facebook posts—are placed in a huge bucket of words that human coders have decided they would like to study. The algorithms not only work to classify words on various dimensions (e.g., positive/negative, level of authenticity, confidence) but can also be used to uncover latent connections between words and collections of words (i.e., phrases).

Geographic Information Systems

A critical element of big data analysis is the layering of data elements to get richer contextual information on environments. From a spatial perspective, leveraging geographic information is critical. Geographic information systems (GIS) deliver various forms of information as multilayered maps in which each layer provides different information. These systems have evolved from complex and difficult applications to essential parts of how citizens and governments understand and relate to their environments. Increasingly, public agencies are using GIS to capture, store, analyze, and manipulate data to decrease costs, increase communication, and improve decision making. The advancements in visualization and information technologies have further accelerated the use of GIS in the public sector in fields such as emergency management and urban planning.

Network Analysis

The exploration of connections helps us understand how networks are structured and how they evolve. Network analysis allows us to see the connections among various data elements. For example, we can see the connections between two agents (humans) or two objects (devices) as well as those that occur between humans and objects (e.g., two humans who are connected to or dependent on the same object—say, an energy source). Today, urban planners must become comfortable dealing with large datasets where computation of networks is not as straightforward as when only a single form of interaction or connection was considered. One of the attributes of big data is the ability to link datasets (i.e., connect data elements) across various domains.

Agent-Based Modeling

In agent-based modeling (ABM), computational models of complex systems are developed by specifying the behavior of agents (individuals or organizations) and their resources, behaviors, preferences, and interactions as well as activity with other agents, objects (artifacts such as buildings, systems, and land), and the environment. Each agent corresponds to a real-world actor who is cognitively and socially bound to a place. The simulation allows one to trace how collective patterns emerge from the interactions between agents and their environments given the rules and constraints modeled. The interaction rules, constraints, and environmental conditions are modeled to represent the options being considered.

Gaming

Planners can also use big data to capture activity, interest, and interaction through gaming. Gaming is a useful and fun method of experimenting and making decisions through gamification, or using game elements in nongame contexts. The point of gaming is to always understand the behaviors of users and what motivates these behaviors. Gaming is a benefit to urban planning because it takes user activity, captured through big data, to understand interactions and decisions. Data collected from games can aggregate such issues as common problems that arise, people's most immediate decisions, what motivates behaviors (e.g., instant gratification, incentives), and what was learned on subsequent tries of the game. Further, gaming can also help transform big data into smart data. This happens through turning user decisions into data-driven insights.

Building Information Modeling

Building information modeling (BIM) is a process of digitally representing the physical and functional characteristics of places such as buildings, tunnels, wastewater and electrical facilities, ports, prisons, and warehouses. BIM is a departure from the 2-D technical drawings that are used in traditional building design. It extends beyond 3-D to a fourth dimension of time and a fifth dimension of cost. BIM provides opportunities to use IoT and other tools, such as GIS, to combine and analyze physical and administrative data (such as vacancies and lease space) and other data sources like LIDAR laser information.

A FRAMEWORK FOR LEVERAGING BIG DATA THROUGH ANALYTICS

The analytical process needed to use big data requires rigor and attention to detail. This process needs to ensure that data are collected from the most appropriate sources, validated, and integrated. In order to do this, appropriate analytical techniques need to be employed, analytical outputs should be carefully scrutinized, and relevant insights and actions should be implemented and disseminated in an optimal manner. The following four-phase framework can help guide big data users through the process.

Phase 1: Management of Data Sources

Data are generated by agents (i.e., humans) and objects (i.e., things). Traditionally, planners were confined to a limited number of sources from which they could collect data. These sources produced data on a regular basis and in defined forms, and they were deemed credible due to their official designations, such as the US Census Bureau, or the regulated processes they followed. Today, things are quite different. In addition to traditional sources, planners have available data from a much wider variety of sources. Many of these sources do not have the same credibility or authority as traditional data sources. They also often lack permanency and emit or produce data that are of varied types, frequencies, and formats. Source management is about knowing what sources one should pay attention to, being able to evaluate the credibility and veracity of sources, extracting data from these sources when needed, organizing sources, and protecting sources and the data being extracted.

Phase 2: Information through Analytics

The next phase starts the analysis process that allows for the extraction of information. When data scientists and analysts receive data, the data will often be in formats that are incorrect, inconsistent, inaccurate, irrelevant, or incomplete. This requires cleaning and correcting of the so-called "dirty data." Data cleaning is not optional, and this task is critical to the analytical process. Without it, insights derived from the data can be incorrect and misguide planning and public policy actions, such as financial or resource allocations or improvements to infrastructure.

Once data from several sources are cleaned, the next step is to link and connect them so as to bring multidimensional perspectives and broaden one's situational awareness. The fusing of data is a nontrivial task that requires creativity and rigor. Creativity is important because identifying the fields that can be linked or connected in a database is not always easy to do, and hence a fair degree of innovation takes place. This is especially true when trying to link databases that do not use the same fields consistently from one database to another. Rigor is important to ensure that the data are not incorrectly manipulated or accidentally transformed during the integration.

A wide assortment of approaches is available to analyze data; several of the common approaches were discussed earlier. While data scientists can conduct analyses, analytic expertise can come from other sources, including community collaborations such as hackathons and crowdsourcing platforms. More traditional analytical tools are available (e.g., software such as SAS and SPSS). Another source are the thriving open-source communities building analytical tools. These tools are made available to the public in the form of source code, technical manuals, and user communities that work on next iterations of these solutions, share feedback, and fix and report bugs.

Phase 3: Interpretation of Data Outputs

This phase is all about ensuring that analytical results are being interpreted in the most accurate manner and within the appropriate context. Analytical outputs can easily be misinterpreted both accidentally and purposefully to advance an agenda. It is important that care be taken so as to not fall into common traps when interpreting analytical outputs. First, interpretation of results in a timely manner is essential. Insights that are outdated and no longer useful are a waste of time for all parties involved. Setting up organizational processes that make data interpretation a regular part of operations is essential to making timely interpretations. Second, the right people should be engaged in the interpretation process. These can be individuals internal to the organization, outside experts, and the public (usually through crowdsourcing activities).

The manner in which analytical outputs are shared is an important decision. Outputs can be shared via reports that describe the results, provide context, and examine the pros and cons of the recommendations or outcomes. Visualizations are a method of sharing analytical outputs using platforms that help create awareness and provide valuable insights into how a city is performing. There are many different visualization types that run the gamut from static visualizations, such as infographics, to dynamic and interactive ones, such as advanced heat maps.

After analytics have been shared, researchers should solicit feedback on the interpretations. This includes critical questions, suggestions, and ideas about future work. It is important to understand the implications of the knowledge in question before constructing a course of action. Feedback can be gathered from citizens, individuals internal to the organization, and outside experts. Citizen feedback can happen in town hall or community forums or through websites with comments sections or citizen engagement platforms.

Phase 4: Design and Implemention of Evidence-Driven Policy

Once the analytics are complete, the process of considering various policy options and designing strategies to implement outcomes begins. When possible, it is advisable to run experiments to test the efficacy of policy options and to study both the intended and unintended consequences of courses of action. Experimentation is a process that planners may not be comfortable with, but it is important to understand how you might go about implementing solutions. Implementation of a solution can take many forms, and often there are trade-offs involved with each possible trajectory. Experimenting might allow you to see how well you witness the intended consequences and to identify some of the unintended consequences. Experimentation is risky, as it might reveal things that warrant going back to the drawing board and beginning again. However, experimentation always leads to learning, which is vital.

When outputs from analysis have been interpreted, shared, and tested through experimentation, they have become actionable insights and reached their potential. At that point, it is then in the hands of decision makers to drive new, evidence-based policies and processes. Policy makers and researchers have recognized that evidence-based policy making based on rigorous evidence is a better way to operate efficiently and strategically. The availability of evidence (using big data and analytics) can help planners understand what is working and what is not.

An interesting option with big data and analytics is to continuously evaluate policies to ensure that the desired implementation and impacts are achieved. It is vital that agencies develop formal processes to evaluate policies on a regular basis. There are several reasons for this. First, conditions change regularly. Any policy that is deployed will need to be revised as conditions in a community change. Second, feedback collected from the evaluation of policies provides opportunities to make small modifications and tweaks. This is much more advisable then waiting until it is necessary to overhaul the entire infrastructure and system.

THE FUTURE OF BIG DATA AND PLANNING

The future of analytics and big data is bright and will transform the practice of planning. While it is always challenging to predict with pinpoint precision how things might play out, there are several trends that warrant attention. First, there will be a growing movement to integrate traditionally disparate databases at the city, state, and regional levels. Today, gaps exist that prevent us from getting a more holistic view of how individuals engage with public services and agencies. Simply put, layering of multiple datasets allows deeper and more precise insights into phenomena. Second, the trend to combine analytics with know-how in terms of behavioral sciences is picking up momentum. Third, the trend of the automation of almost everything is in full swing. Automation will also affect how data are collected and who analyzes the data. Fourth, given how much people's lives, organizations, and societies have come to depend on the digital information infrastructure, it is not surprising that the unscrupulous want to disrupt and cause harm to these systems. In the future, municipalities and different government agencies will need to find ways to coexist; for instance, officials concerned with big data and other technologies must learn to function effectively with officials concerned with privacy and security. Fifth, the rise of the nontraditional planner and the growth of crowdsourcing platforms mean that we are likely to see many more solutions developed outside of local government.

Urban planners play an important role in most cities as they create and implement visions for communities. They lead efforts around social change, policy making, community empowerment, sustainable growth, innovation, disaster preparedness, economic development, and resilience. Planners also solve problems through research, development, and design. Fundamental to carrying out these activities are their abilities to make decisions in effective and efficient manners. To increase their decision-making capacities, it is critical for planners to leverage data in innovative ways. The future offers exciting possibilities, and planners are encouraged to stay abreast of developments in data science and to find ways to begin conversations with city leadership about how these technology-based opportunities can advance their communities.

CHAPTER 1

INTRODUCTION

Over the decades—and centuries—planners' roles have differed and progressed with the challenges and needs of communities. Planners have dealt with far-ranging issues such as the growth of technology, industrialization, shifts in ethnic makeup, natural disasters, war, poverty, and a host of other occurrences that have directed their urban planning efforts. Today, planners have a new issue to contend with: big data. Big data involves datasets that are larger and more complex and that exceed the capabilities of traditional data processing applications. Unlike other issues, big data is not a problem but an opportunity.

We have more data than ever before on all kinds of things planners care about, ranging from fixed objects, such as buildings, to dynamic activities, such as driving behavior, and at all levels of granularity, from cities to individuals and everything in between. In addition, we have unprecedented access to data in real time, which gives us new capacities to monitor and evaluate the state of objects (artifacts) and agents. We have technology to thank for a great deal of this. Much of what we interact with is fitted with technology sensors, and our own (consumer) technologies—such as mobile phones, tablets, and fitness trackers—are also valuable sensors and emitters of data.

Before there was big data, there were data. Organizations have always generated and collected data, but the data were often stored in a way that was not useable, not made publicly available, or not leveraged correctly to gather any meaningful insights. Additionally, data quality and up-to-dateness was a common problem. Traditionally, planners have relied on US Census American Community Survey data or geographic information system (GIS) shapefiles that in many cases were not regularly updated. Another critical reason for the inefficient use of data was that the tools were not available to do the analysis or collection. In essence, there were few planners who were getting the best out of their data in terms of using it to help arrive at decisions.

Now, we have big data. As recently as the year 2000, only one-quarter of the world's stored information was digital (Cukier and Mayer-Schoenberger 2013). The Internet has changed much of how we function and has given us more data than we know what to do with. According to estimates, the digital universe of information will double every two years through 2020 (Duetscher 2014). Further, more advanced and affordable analytic tools are helping us leverage data to develop smarter cities that are cost efficient and innovative and make data-driven decisions that will improve the lives of citizens.

Information technology is transforming how we think about planning and the future of our communities, even when it comes to smaller, traditional datasets. For example, in Mobile, Alabama, code enforcement officers are fighting blight using mobile phones and Instagram (Wogan 2015). Over two weeks, 14 code enforcement officers recorded 926 blighted properties in the city. This gave officials a better understanding of their blight problem and a place to begin in solving the problem by exploring why this occurs and what can be done about changing those factors over time.

Social media has grown beyond "likes" and "retweets" to actionable information. Information exchange today is global and instantaneous thanks to technologies like smartphones, tablets, and computers. The masses of data that social media platforms produce daily are unfathomable, but the data are not valuable if they are not properly used or understood. We are seeing more sites and apps providing analytics to help users understand page visits, views, and popular content. We saw new trends in the 2016 presidential elections where candidates were tweeting and snapchatting to reach millennials and achieve greater impact. The pervasiveness of social media means that it is heavily used and information moves quickly. This does not mean that the information is sound, factual, or even relevant. But because information travels so quickly, it can begin to take on a life of its own. Consider your favorite meme or viral video and how you stumbled across it. Later in this report, we will explore how social data can be used for actionable insights.

The power of big data is exciting as well as frightening. Big data combined with a little inattention or ill intentions can have disastrous outcomes for the public. Following a tornado that left neighborhoods ravaged in Texas, demolition crews using Google Maps set out to demolish damaged properties. Google Maps uses a combination of aerial, satellite, and street-level images and data to create its detailed maps of the world—but, as previously noted, just because data are available does not always mean those data are correct or up to date. The demolition crew used Google Maps to get directions to the house in Rowlett, Texas, that they were supposed to demolish but were led to a house a block away from the actual location. The crew did not confirm the street they were on and proceeded to level the wrong home (King 2016).

Behavioral economists are using big data to "nudge" citizens toward positive behaviors such as paying taxes on time, being environmentally conscious, or donating to charity. Studying the economics of how people make decisions and how the provision of incentives and information change decisions and preferences is at the core of behavioral economics. The UK Behavioural Insights Team (also known as the "Nudge Unit") (www.behaviouralinsights.co.uk), a limited liability company that was initially set up in the UK government's Cabinet Office, conducts studies on nudging in the public sector to enhance outcomes. It encourages citizens through text messages to give a day's salary to charity or pay fines. Analytics has assisted its work by offering more targeted nudges. For instance, the Nudge Unit sends letters and text messages to citizens reminding them to donate or pay a fine. The notices use different levers to obtain citizen compliance, such as applying social pressure by letting them know that their neighbors have already done so. The Nudge Unit is able to personalize its communications by adding names and using language directed at different types of professionals (e.g., doctors, laborers), further increasing compliance.

Technology is not only generating these data but is providing us with the computational tools to analyze, visualize, interpret, and act on large amounts of data. In addition, we have tools that are automating decision making through the embedding of intelligence, which is derived through analyses of large amounts of data. In 2014, UPS implemented a new driving strategy to enhance driver efficiency and safety and lower emissions; the plan was to eliminate left turns in drivers' routes. Left turns across traffic can be dangerous, and waiting to make a left turn can delay drivers, while they can always turn right on red lights. UPS (2016) did this through a proprietary GPS system called Orion, which functions like regular GPS and has the added capability of specifying package delivery during a certain window of time. Orion's calculations of the most efficient travel paths resulted in routes that almost never included left turns. Through the visualization process, UPS was able to see how this increased efficiency manifested itself through the delivery process.

Despite these new computational tools, leveraging data is not a new concept for planners. Planners have needed and always will need to collect, validate, and analyze data. In the 1960s and 1970s, the City of Los Angeles used data to manage and improve problems in high-poverty urban areas (Vallianatos 2015). These areas were largely unsafe and unsanitary, and the living conditions created a number of social, health, and economic challenges for residents. Los Angeles urban planners used data to help them understand the problem before defining a solution. First, they hypothesized that these living conditions threatened health, social, and economic prosperity. To test this hypothesis, they used data from a variety of sources to develop cluster analyses of neighborhood demographics and housing quality. Information from those analyses (such as prevalence of vacant homes, street conditions, land use, and blight) was used to direct resources to address crime, unemployment, street maintenance, and traffic and to locate community amenities such as libraries and parks in low-income areas.

Today, using data to understand populations is the norm, but now the stakes have been raised; data are being leveraged to do more and to make cities smarter and more intelligent. The central goal is the ability to collect and analyze data, make predictions, build scenarios, and act on the data in real time to ensure that cities are responsive, sustainable, and resilient. An example of a smart-city innovation is an initiative developed by District Department of Transportation (DDOT) in Washington, DC, to address bus bunching (Eagleburger 2015). Bus bunching occurs when two or more buses scheduled to run the same route at evenly spaced intervals instead end up arriving at the same stops at the same time. This causes unreliable service and results in overcrowded or empty buses. DDOT partnered with EastBanc to install smartphones on circulator line buses. The smartphones transmitted bus position signals every three seconds, which allowed DDOT to alert bus drivers in real time to speed up or slow down to avoid bunching. Using real-time tracking data, DDOT was able to modify routes and maintain bus arrivals at 10-minute intervals.

Cities, businesses, universities, and community organizations all have the potential to leverage data and

BIG DATA AND THE DIGITAL DIVIDE

Technology has greatly affected attitudes and lifestyles. The Pew Research Center reported in 2015 that 66 percent of Americans owned at least two digital devices—laptops, smartphones, or tablets—while 36 percent of Americans owned all three (Anderson 2015). It also found that nearly two-thirds of American adults, 65 percent, use social networks, with 90 percent of young people (ages 18–29) using social media. The rise of social media has modified Americans' daily behaviors that relate to civic life, parenting, dating, news consumption, communication patterns, politics, and work (Perrin 2015). The way many individuals receive and disseminate information today is largely technology based.

But not everyone is part of the technology revolution. The term "digital divide" refers to the gap between groups of individuals with access to modern technology, such as digital devices and the Internet, and groups such as lower-income individuals and the elderly that do not have such access to or cannot afford digital services. The digital divide has big data implications when it comes to urban planning. Obviously, it is difficult to collect data on individuals who are not generating data in more common ways, such as through Internet usage, social media usage, or credit card transactions.

According to the Pew Research Center, 67 percent of Americans have broadband Internet at home (Horrigan and Duggan 2015). Various states, such as Missouri and Ohio, have sought to bridge the digital divide through efforts to install fiber optic networks that provide Internet access. Outside of the United States, digital penetration can be well below the US national average in particular areas and social groups for different reasons. In Europe, the digital divide is widening because of aging populations. In developing countries, digital penetration is lower in rural areas, which are the places where most citizens reside (Internet World Stats 2016).

Urban planners must be mindful of this issue when using big data to chart the course of their cities. Bridging the digital divide is necessary to seize all of the opportunities the Internet and technology offer. These opportunities include the ability of citizens to acquire new skills and be better informed and of cities to develop more comprehensive sets of data that include the majority of the population and not just the few privileged enough to have access to technology.

THE MOBILE EXPLOSION

Mobile phones are a source of much of the data being captured. Call data, text data, Internet usage, app usage, and telephone metadata (e.g., location, time of call, duration of call) create significant opportunities to leverage data to create new insights. For instance, scientists from the Data Science Lab at the Warwick Business School used mobile phones to estimate crowd sizes (Warwick Business School 2015). They analyzed the volume of incoming and outgoing calls and text messages and the number of active Internet connections to determine crowd sizes. They found that spikes in use coincided with football matches and accurately corresponded to the number of people attending the match.

Further, mobile phones are windows to another world for many in developing countries. Mobile phones are being used to assist farmers, increase secure banking opportunities, slow the spread of disease through education and assistance, and conduct research. In some cases, big data and analytics are an alternative to poor record keeping in developing countries. Researchers at the University of Washington and the University of California, Berkeley, analyzed billions of mobile data points from Rwanda's largest cell phone network to provide current information about the socioeconomic status of citizens (which is not done regularly or accurately in the country). The study captured information about users' travel patterns, social networks, and patterns of communication (Choi 2015).

In addition, the researchers conducted surveys with 850 phone users about their assets and access to assets—such as quality of housing, car or motorcycle ownership, and other indicators of wealth and poverty. Based on the asset data and usage patterns, they were able to develop accurate maps about the poverty and wealth characteristics of users (Blumenstock, Cadamuro, and On 2015).

analytics to improve outcomes with the unleashing of numerous new technologies and greater data capabilities. Perhaps the most significant and substantial data capability that has grown recently is this proliferation of big data and analytics. Access to big data enables us to do more insightful analysis that was previously not possible. For example, we can do more real-time analyses of our environments, and we can also do analyses that combine data from multiple subsystems within environments to get a more holistic picture and broader understanding. We can also mine data from past activities to discover latent patterns and develop predictive algorithms.

Partnerships will be essential for leveraging big data in the future, and data collectives are one way to promote the sharing of data (both large and small datasets) to drive solutions for more complex problems. Back in 1990, a *Harvard Business Review* article entitled "Information Partnerships—Shared Data, Shared Scale" stated that "information technology empowers companies to compete, ironically, by allowing them new ways to cooperate" (Konsynski and McFarlan 1990, 114). An important form of cooperation is public-private partnerships—for example, when airlines work with law enforcement on public safety or private satellite companies offer data to contribute to national weather accuracy.

Early examples of data-sharing partnerships are shaping planning. Rideshare companies Uber and Lyft are sharing data on their services in select cities to help transportation officials understand how they are modifying traffic and transit patterns. This will be important as these types of services become more mainstream. For instance, the Philadelphia Union, a US soccer team, partnered with Uber, which has a dedicated pickup and drop-off location at the stadium (George 2016). Deals of this sort are likely to modify how people assess their transportation options and decide to travel. But to understand these patterns and trends, data are needed. This PAS Report offers useful, reasonable explanations and methods to help urban planners move toward developing smarter cities. The report focuses on helping planners leverage the potential of big data through a growing understanding of data management, analytical tools, intelligence management, and strategic management that both transforms and embraces the complexity of cities.

PLANNING IN THE ERA OF BIG DATA

Cities are changing rapidly. Globalization has increased our reach to the world and the reach of others to us. Cities are emerging (e.g., cities in China) and fading (e.g., shrinking Rust Belt cities in the United States) in much shorter timeframes. Issues such as refugee crises and food shortages are forcing cities to confront challenging questions on all fronts. This will not change but will rather become more persistent as technology lowers the barriers between cities. Thus, planners' use of big data will not be purely elective.

Expectations of planners will change and requirements such as more efficient use of resources will become more prevalent. Customers and elected officials expect information and action quickly. Processes that require lots of time to conduct will no longer be acceptable, and rightly so. Rich data harvested from digital technologies, the Internet, and social media will be used to improve conditions and the future of communities.

In his TED Talk "The Birth of a Word," MIT professor Deb Roy (2011) explained that as the world develops the capability to gather data and connect dots in ways that were not previously possible, we are able to see new structures and dynamics. As a result, we can understand human behavior and society in different and more personal ways (versus at the city or census tract level). This new ability opens our world to possibilities never before conceived of because the insights have never before been experienced. We are approaching a new frontier of massive data flows and limitless computational power that is allowing us to be increasingly self-aware, connected, and, as a result, able to make profound transformations.

However, these transformations do not happen all at once, and different phenomena are influencing these developments. First, we are seeing small changes that, when taken together, have increased societal well-being. In an era when metropolitan cities are so significant to national growth, smart-city growth actually becomes growth for the entire country. According to the Brookings Institution, there are 388 metropolitan areas in the United States that account for 84 percent of the nation's population and 91 percent of the gross domestic product (GDP), with the 100 largest metropolitan areas generating 75 percent of GDP (Kulkarni and Fikri 2015).

In London, the population is at 8.6 million and is expected to grow to 10 million by 2030 (BBC News 2015). The Greater London Authority (2015) expects 641,000 more jobs, 800,000 more homes, and 600,000 more passengers traveling at peak times by then. Technology and innovation are key methods of accommodating the population influx as well as reducing traveler stress. In 2003, the Oyster card, a form of electronic ticketing the size of a credit card, was issued for the transit system in greater London. It was designed to reduce the number of ticket offices and paper tickets. Prior to the Oyster card, the cost of maintaining ticketing offices and producing tickets was nearly 15 percent of the cost of transit operations in London (Ahmed 2015). Today, the use of and enhancements to the Oyster card over its 13-year existence has resulted in a decrease in the cost of ticketing to slightly lower than 10 percent. With the availability of contactless debit and credit cards such as Apple Pay or Samsung Pay, the cost is expected to drop to 7.5 percent. The incremental reduction in costs might seem minor, but the savings means that more resources can be used to enhance transit services, reduce transit fares, and decrease passenger frustration—all of which will enhance city functions. A 2013 study by Accenture about what passengers want from transit providers revealed that travelers still want the features technology offers, such as ticketless access via smartphones. They are also willing to pay more for the convenience, and they want to be connected to the transit system through social media to have accurate information at all times (Accenture 2013).

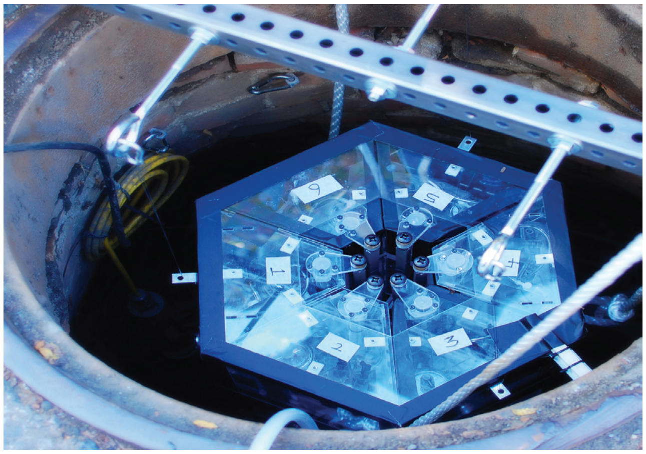


Figure 1.1. Mario is the first-generation sewage sampler developed for use in the Underworlds project (MIT Senseable City Laboratory)

Figure 1.2. Researchers operating Mario in Cambridge, Massachussetts (MIT Senseable City Laboratory)

Second, we are living in a time of unprecedented access to data on just about anything you can imagine, and there is no indication that this trend is going to slow down. Sensors on our bodies (e.g., Fitbit, Garmin Vivofit, Jawbone), the devices we carry (e.g., mobile phones), and technology embedded within physical environments (e.g., traffic cameras) and virtual environments (e.g., social networks) collect data on a constant basis. The volume of data that is being created is exploding and it will only grow in the years to come. By 2020, 1.7 megabytes of new information will be generated per second for every human on earth. Of that data, 40 percent will be stored in the cloud. By the time this happens, our digital universe will have accumulated 44 zettabytes (or 44,000,000,000,000,000,000,000 bytes) of data; today, we have around 4.4 zettabytes (Adshead 2014). Further, two-thirds of Internet traffic during this time will be generated from devices other than personal computers (Cisco 2015). Despite the high volume of data being generated currently, it has been estimated that only 0.5 percent of those data is being analyzed. For urban planners, this explosion of data creates a new world of opportunity.

Digitization of all aspects of the city—from sensors embedded in our infrastructure to activity and movement tracking capabilities through GPS and even social sensing (crowdsourced data from sensors closely attached to humans, such as mobile phones or wearables)—allows us to learn about the sentiments of residents regarding policy and social options and provides a distinct opportunity to rethink how we plan and manage cities and communities. Key to this is the ability to leverage data through analytics in an effective and efficient manner. Real-time analytics applied to data are also increasing our situational awareness about our environments. This increased real-time situational awareness will enable individuals, communities, and organizations to make more intelligent, or smarter, decisions.

For instance, researchers from MIT and Kuwait have partnered to study epidemiology in real time through the Underworlds project (http://underworlds.mit.edu), which is creating a smart sewage infrastructure in Kuwait (Figure 1.1 and Figure 1.2) (Kuwait–MIT Center for Natural Resources and the Environment 2015). Underworlds embeds biomarkers in municipal wastewater and sewage networks

BIG DATA AND THE PLANNING PROCESS

Nader Afzalan, Chair, Technology Division, American Planning Association and Visiting Assistant Professor, University of Redlands

The invention and adoption of new technologies and online sources has facilitated the creation of big data. These large quantities of data can greatly advance or hinder planning processes. Big data promises novel opportunities for planners to explore various urban and environmental systems and to make more accurate predictions. However, it can also mislead planning processes if not used appropriately to address planning issues—for example, using big data to respond to topics of public interest.

Planners use various types of data and data sources in their complex plan-making activities. Big data has the strong potential to enrich various stages of plan making, including visioning, problem assessment, scenario planning, and plan implementation. Big data empowers planners and decision makers by helping them to better understand current situations and predict future ones more precisely and accurately. The intense amounts of time and labor required for data collection as part of comprehensive planning might be resolved in the near future with big data generated using volunteer sources or created through more formal processes. However, planners will need to be prepared to deal not only with data quality issues but also the institutional challenges of collecting, analyzing, and incorporating big data into planning processes.

In terms of organizational capacity, planners need to consider a range of issues. Does the planning organization have the staff or financial capacity needed to collect, organize, and interpret big data? This is particularly an issue for local and regional governments. Does the organization know about and understand the potential privacy and legal concerns of using big data generated from online sources in its planning process? Does it have the capability to collaborate with outside organizations in order to most effectively use big data?

Planning organizations have limited resources to make plans and decisions within a certain period of time and in complex political and technical environments. Planners should evaluate the technical and political challenges of planning processes in order to use big data to make more precise and equitable plans and decisions. If the organization is not able to use big data effectively, planners should think about either collaborating with outside organizations or using traditional data sources.

While the proper use of big data by planning organizations can greatly enhance the quality of plans, its inaccurate use may result in plans that are based on invalid or biased assumptions. For example, big data collected through parking sensors can help transportation planners conduct more precise analyses or develop better predictions of traffic flow and parking demand. On the other hand, if these data are used as the only data source, the result may be a plan that does not represent the real needs of the community.

While big data can help planners respond to the complexities of planning processes, it can also add another layer of complexity. Big data can make our cities smarter, but it can also make cities data obese. Planners should be ready to employ these new data sources in their planning work based on their organization's capacity to use big data and the effectiveness of big data in responding to issues of public interest.

PLANNERS ARE NOT IN IT ALONE

Foundations and private companies have stepped in and shown their interest in city issues through funding and the provision of resources. Bloomberg Philanthropies (www.bloomberg.org/program/government-innovation) has been extremely active in the space of government innovation through several grant-funded programs that help cities find solutions to problems they have identified. Similarly, the Knight Foundation works with local governments to tackle problems through the Knight Cities Challenge (www.knightcities.org) innovation program.

The private sector also supports government innovation through public-private partnerships. Google's Sidewalk Lab is a partnership with the US Department of Transportation that seeks to provide technological solutions to urban problems, most specifically transit problems. This is happening through a digital platform called Flow (www.flowmobility.io) that will be made available to cities around the country. Flow will aggregate and analyze mobility data from a number of sources such as Google Maps, Waze, and municipal data. This initiative will seek to address current transportation problems as well as help with future decision making on emerging issues such as the use of autonomous vehicles (see "Autonomous Vehicles," p. 40).

A partnership between MasterCard and Cubic's transportation systems consultancy service, Urban Insights, produced the Urbanomics Mobility Project, which aims to help governments make better decisions based on data (Silver 2015). By using billions of pieces of purchase data, the project looks to find relationships between how people travel, what they buy, and how this information can build better cities. The goal of the partnership is to build a useful tool for an array of stakeholders—such as urban planners, commercial real estate developers, and city and regional government officials—to help them manage efficiency.

The project helps shape decisions by providing governments with data about economic activity in the state and region by market. The project can model transaction data to find gaps in public services, transit, shopping habits, home location, and job growth in cities. Data about home location might reveal a concentration of more young people in city centers. This information could help with economic development that could focus on bringing targeted stores and nightlife into the area and developing civic spaces. This project will begin deployment in 2016 (MasterCard 2015).

in Kuwait to collect spatiotemporal sampling data and develop real-time public health profiles of urban areas. The project's researchers aim to track antibiotic gene resistance and multi-drug-resistant bacteria, and identify other health and biosecurity indicators—all through combining big data and innovative analytic programs. One of the first of its kind, Underworlds is proving that wastewater systems can be data-valuable sources for public health and prevention, providing an alternative to running costly tests on individuals. Further, the project is serving as proof of the value and advances that can be arrived at through innovation, data, and technology.

In another example, analytics was used to improve air quality. Air quality is a concern for in many industrialized nations and cities. However, cities have a difficult time understanding air pollution patterns because satellites that measure particulate matter largely generate baseline air quality data but do not track real-time changes in cities' levels of air pollution. The Internet of Things technology, big data, and analytics have given rise to new opportunities to measure air quality using small, portable air sensors that collect data in real time. The environmental research technology startup Aclima used its AirBeam air sensors to measure and understand daily air pollution levels within cities (Spector 2015).

Aclima installed 6,000 sensors in 21 Google buildings and on three Google Street View cars that were calibrated with sensors from the US Environmental Protection Agency. The presence of sensors at all times and the collection of real-time data enabled Aclima to better understand changes in air quality and to pinpoint reasons for these changes. For instance, during a two-day strike of the Bay Area Rapid Transit (BART) system in the San Francisco Bay Area in 2013, more cars were back on the road. The sensors revealed that levels of particulate matter increased inside of buildings during this time, as smog can have higher concentrations indoors compared to outdoors. Conversely, on Bike to Work Day, Aclima found less particulate matter seeping into buildings. By comparing these trends, it was able to estimate the amount of pollution that BART prevents through its regular operation.

However, there is still much work to be done in this area. Aside from large cities like Chicago, Philadelphia, New York, and San Francisco, local governments (especially smaller ones) are struggling with what constitutes meaningful use of big data and how to apply analytics to planning and management functions. In most cases, these technologies were not designed for adoption by smaller communities but are geared for big cities. Issues surrounding investment, internal expertise on issues, governance, and organizational goals all are considerable factors related to the uptake of big data.

Also, big cities are better able to leverage technology due to the volume of startups that are located in these areas, which creates learning opportunities and symbiosis between government technology and the greater technology community (Renn 2015). Professional organizations have tried to help local governments find their way in this area. For instance, the International City/County Management Association (2016) offers low-cost resources for smaller cities, but most efforts have not taken off as would be expected. For this reason, more solutions need to be developed that are light enough, but also meaningful enough, for small communities to adopt.

Third, our physical environment also presents us with unprecedented challenges. Urbanization is happening at a rapid rate and this can leave our cities in untenable positions. In developing countries, urbanization trends have a particularly challenging effect on urban planning. On average, the rate of urbanization in the developing world is five times the rate of the developed world. Cities in the throes of urbanization are likely far from resilient and are, in fact, likely unstable and fragile. Fragile cities are cities that are likely to crumble when faced with the severe stresses of crises or disruptions, which can result in deaths, severe damage, and battered economies.

Further, rapid increases in urban population density can create environmental problems that exacerbate impacts to infrastructure, health, social systems, and climate change. Planners and citizens in these cities are confronting these challenges by using resources such as local knowledge and skills to meet needs and plan for the future. At the same time, the need for planners to leverage data and information for designing smarter and sustainable environments is at an all-time high, with enhanced decision making presenting itself as a leading strategy for managing the next waves of urbanization.

APPRECIATING THE COMPLEXITY IN PLANNING

Cities are complex ecosystems that are characterized by emergent properties wherein the whole is greater than the sum of its parts. This complexity means that the ecosystem can (and often does) behave in nonlinear ways. Factors such as history, sudden change, legacies, and unpredictability are significant features of urban ecosystems. This complexity will only increase as more elements are introduced; smart-city tools such as analytics and algorithms help to moderate the complexity and provide new ways to think about urban systems from a perspective where complexity is appreciated rather than shunned.

The complex nature of cities can be found in the interdependencies within these systems. These inter-dependencies comprise different sectors (e.g., energy, banking, transportation, water, agriculture) that are publicly and privately owned as well as communities that coalesce to make an interconnected system that must be secure to carry out its services. Inevitability, this interconnectedness will be exploited by human decisions and human-made disasters, which creates a significant need to understand and manage the complexity adequately.

Given the complex nature of cities, silo-based planning efforts are not going to work. Big data allows us opportunities to move beyond silos and use the complexity of the city to our benefit. It enables the integration of data from multiple systems and environments so that we can analyze behaviors and study the impact of plans, both in terms of intended and unintended consequences, more holistically. This allows us to know if a plan or policy is effective across an entire city and to identify its effects at the regional, state, and even global levels.

The role of the planner is also changing. Previously, there were designated departments or organizations to plan cities; today there is also an emergent community of civic innovators who are building solutions that are in the realm of planning. These solutions are more often than not ones that take advantage of data in novel ways, so when it comes to the changing nature of cities, we have new sources of knowledge and participation that need to be leveraged. As a result, rather than planners trying to design for a city or community, they need to embrace designing with these groups, where the medium for engagement is normally around data and data-driven solutions.

In the 1960s, urbanist and journalist Jane Jacobs (1961) wrote in *The Death and Life of Great American Cities* that the decline of cities could be attributed to urban planning and urban policy. Jacobs noted that the problem with urban planning was that planners did not acknowledge that individuals lived in communities characterized by layers of complexity. Planners viewed cities as buildings and people and did not recognize their emergent properties such as healthcare systems, financial markets, families, job markets, and urban communities exhibiting rational and irrational behaviors. As a result, planners took a singular approach to urban planning by looking to control a few specific aspects of urban systems and using quantitative methodologies to guide planning. This was most evident during the US urban renewal period, where one aspect of city life—land redevelopment in high-density areas—was pinpointed as needing improvement; the result was inattention to the many other networks and elements at play that caused suffering and displacement in communities. Instead of seeing the city singularly, Jacobs posited that planners must observe the city as a whole and then make plans.

Expanding on Jacobs' notion of seeing the city as a whole of many parts, Batty (2012) explained that cities are environments open to the wider world, ever-changing, far from balanced, and evolving from bottom-up decisions of individuals and groups with the occasional top-down centralized action. He asserts that, as a result, cities are complex, living organisms that can only be understood through real-time monitoring and analytics; this insight helps us to assess their current status and to make decisions for the future (DeAngelis 2014). Until recently, real-time monitoring of cities was impossible. However, the ability to monitor and analyze data in real time means that planners are able to increase the efficiency and capacity of cities by connecting, integrating, and analyzing information.

Today, planners face the complex task of blending policy, community engagement, sustainability, economic viability, emerging technologies, and more scrutiny than ever to effectively plan for their communities. But this complexity and interdependency creates fertile ground for cascading failures throughout a system. A cascading failure is a failure in one part of an interconnected system that when triggered starts failures in other parts of the system. As failures can cascade through a system, so can early warning signs of problems.

Early warning signs are symptoms of a problem, such as higher than average levels of property abandonment in the case of blight and lower than average levels of maintenance in the case of public infrastructure. These symptoms are not to be confused with the causes of failure, which can be lack of management, complacency, or inattention to issues. That is the benefit of complex systems; the problem will show up before the failure.

Now, with smart technologies, there are greater opportunities to become aware of early warning signals through real-time access to sensors and predictive models that alert us to problems. The use of data, analytics, and smart technologies will help cities manage their complexity and become less reactive to challenges and threats.

THIS REPORT

This report was developed to help urban planners further understand big data and analytics and offers real-life examples and advice for smart-city development. Chapter 2 will discuss in detail what exactly big data is and describe how big data coupled with analytic tools creates new opportunities for cities. Chapter 3 describes popular analytic tools along with examples of how these tools are currently being used in the field. Chapter 4 lays out a framework for leveraging big data and analytics and discusses four essential phases of management to correctly and adequately use the insights derived from big data analyses. Chapter 5 concludes with a discussion of the future of big data and analytics and important concepts to bear in mind as technologies proliferate.

FUTURE: ALGORITHMIC REGULATION

Imagine a world so connected and self-sustaining that we could get rid of most levels of bureaucracy and allow algorithms to help enforce laws. That is the basis of algorithmic regulation: Tim O'Reilly's concept of an alternative form of government that utilizes algorithms to consistently enforce regulations and law enforcement (Araya 2015). This is different than the discordant efforts of the many bureaucracies that are hired to enforce laws but do so unevenly because of power differentials, poor funding, or poor enforcement. Algorithmic regulation would do all of the enforcement and do it consistently. In planning, technology growth is advancing at a breakneck pace and will be a heavy burden to bear. Soon, for instance, autonomous vehicles will be a reality. We can combine algorithmic regulation with the growth of autonomous vehicles to ensure safety and observance of the law. Although it might seem a little far-fetched at this point, technology is transforming everything and innovative responses are needed to manage the change.

EARLY WARNING SIGNS

In 2014 the metropolitan region of Atlanta experienced a disruption in its transportation system due to a snowstorm and an already fragile highway system. The storm created a traffic jam that resulted in more than 2,000 school children spending the night in buses, police stations, or classrooms and trapped people in their cars for up to 20 hours. Total snow fall? Just over two inches. The lack of extraordinary circumstances that was able to cripple one of the largest metropolitan areas in the United States was a wakeup call about the fragility of the city's travel infrastructure.

On Tuesday, January 27th, between noon and 5:00 p.m., schools and the majority of businesses released their occupants to return home to prepare for snowfall. Nearly a million individuals that worked in the city of Atlanta headed towards the "Downtown Connector" highway to either collect their children from school or make their way home. By nightfall, individuals were trapped inside of their cars or indoors due to traffic congestion. This traffic congestion underscored decades of poor transportation management that culminated in the mass traffic jam.

In 2000 the Brookings Institution noted that much of Atlanta's traffic congestion was due to unbalanced growth patterns. These growth patterns, a result of suburban sprawl, stemmed from job growth. The fastest-growing and most lucrative job sectors were located in the northern part of the region. Many of the upwardly mobile residents of the Atlanta metropolitan area lived in the suburbs and commuted by car daily to the northern region of the city (Brookings 2000). While Atlanta does have a public transportation system, it is disproportionally frequented by nonwhites and less-affluent riders, only services two counties, and predominately serves the southern, poorer part of the metropolitan area.

The hypergrowth that has taken place in the Atlanta metropolitan area resulted in an unbalanced region and a vulnerable transportation infrastructure. Fast and aggressive growth in the northern and outer portions of Atlanta and slow growth in the southern portion had a significant effect on the worsening of traffic and the ability for a medium-sized disruption to wreak havoc on the city. Big data and analytics can help with this by helping planners understand system weaknesses and plan accordingly when early warning signs arise.

TOP DOWN, BOTTOM UP, OR SOMEWHERE IN BETWEEN?

In their efforts to help their communities become smart cities, planners have a serious decision to make: whether they are committed to a top-down, bottom-up, or somewhere-in-between approach. Top-down approaches that are driven by planners are likely to have financial backing from local leaders, but if not managed correctly they can miss a critical element of the smart city: the people. Songdo City in Seoul, Korea, is an ambitious smart city built from scratch. Designed to become a highly networked city of the future, Songdo City was extremely popular with urban planners, but there have been serious problems attracting businesses and people to the area (Falk 2013). Planners had the resources to build the city they wanted but lacked significant buy-in from the individuals and groups that mattered most.

Bottom-up approaches are much more citizen-directed, which requires the relinquishing of authority on the part of planners. With this approach, citizens are able to lead smart-city development through engaging with open data, working with their local leaders, and developing their own innovations, such as building their own platforms to provide citizens with information. For instance, LocalData (http://localdata.com) is a digital platform that began as a Code for America project in Detroit. It provides urban planners (and others) with the ability to gather real-time data on project-specific indicators (e.g., community assets, property conditions, building inspections) and then visualizes the data.

Both approaches have merit, but success really happens when top-down and bottom-up unite to transform a city. Amsterdam is an exemplar in this capacity; the city put together a "smart city platform" that allowed citizens and businesses to develop and test green projects. The city's IoT Living Lab (https://amsterdamsmartcity.com/projects/iot-living-lab) puts citizens, government, and businesses to work by testing new products and ideas in real-life settings. It allows everyone to develop solutions and does not close feedback loops or opportunities for implementation. In some areas of Amsterdam, the city utilized participatory service design where, for example, schools are rated by crowd-judging based on student and parent input.

CHAPTER 2

BIG DATA AND ANALYTICS: THE BASICS

The availability of big data and its associated tools offer planners the ability to measure aspects of the city at finer scales, study more variables, and make even more connections between the built and natural environments (Glaeser et al. 2015). Given the value and interest in big data, we should be surprised that there still is not a widely accepted definition of this term; most commentators will agree that they "will know it when they see it." This chapter will cover some of the elementary concepts that planners need to be aware of when exploiting big data through analytics.

WHAT IS BIG DATA?

The simple answer is that big data is data that are, for the lack of a better and less clichéd word, big. A set of data is big when it is too large and too complex to be stored, transferred, shared, curated, queried, and analyzed by traditional processing applications. There is no specific size that is assigned to big data, as it is always growing.

A more complex and accurate answer requires the understanding that data can be big in several dimensions. The simplest dimension is volume. We are talking about data that ranges in size from terabytes (1000^4) to petabytes (1000^5), exabytes (1000^6), and zettabytes (1000^7). With advances in computation, what is considered big today will be small in the near future, and what is gigantic today will be medium-sized in the short term.

The velocity at which data are generated and collected has undergone major changes over the last few years. Today we generate data at an accelerated rate across a multitude of environments. The lag between when an activity is conducted and when it is recorded has, for most things, almost disappeared.

The variety of data streams, formats, and types of data that we have access to has also exploded. Today, we have to contend with data that come across multiple streams—from formal systems to informal platforms (e.g., social media)—and formats including text and audio, visual and aural, and even olfactory. These data streams emit data with a high degree of variability in structure, frequency, predictability, and other characteristics.

Given these realities, ensuring veracity and validity of data is essential. The pressure to act on data in real time only obfuscates further the issue of ensuring data accuracy and credibility of sources. For transient or sporadic sources that enter and leave our environments, traditional methods to evaluate accuracy and credibility do not work well. Innovations are needed that take advantage of collective intelligence and crowdsourcing models. Transient sources do not have performance histories and so we have limited information on how to gauge their credibility and reliability. This is akin to new firms or products that enter a marketplace and start with no credibility. Only after they receive reviews and positive ratings is that credibility established. There are new data sources that pop up frequently that purport some sort of value, but their value needs to be accounted for.

The value that one gets from big data has the potential to be high if the data are analyzed in an effective and efficient manner. One of the major drivers of the value of big data is the ability to link datasets spatially and temporally. Linking datasets across different systems and sources uses data that have not been previously linked, which enables new functionality and can result in richer and broader situational awareness. We can situate, analyze, and visualize the data spatially on traditional layouts such as geographic maps. From a temporal perspective, we can see how various variables change over time. We can also view variables at different levels of granularity to get at the big picture and drill down to see specific details at very fundamental levels.

The demands of growing urban areas necessitate the use of behavioral patterns derived from individuals' cyber footprints to create predictable, evidence-based assessments of the future; data are key. Big data and access to analytics offer solutions to what Luis Bettencourt (2013) refers to as the planner's problems in finding the best social and spatial configuration of the city: (1) knowledge and (2) calculation. The knowledge problem is the lack of knowledge (or data) available to planners about the state of their systems (cities). The calculation problem refers to the computational complexity—the number of steps necessary to evaluate a scenario—of performing planning tasks. The knowledge problem is increasingly diminishing due to the increase in technological capabilities and storage capacities for collecting lots of data, while the calculation problem can grow more complex with the changing volume and variation of data and data quality rules.

Big data also allows for the linking of datasets. This aspect of big data reflects its variety; datasets in different formats and from multiple sources can be joined together. For example, data from people and machines can be merged and analyzed together. This is a key feature of smart cities and how data adds value. For instance, in Tucson, Arizona, a 10 year-old girl was abducted from a playground in front of two other children and no adults. The police were able to use predictive software that included information from a host of datasets—such as arrest records, accomplices, known associates, car registrations, known addresses, and aliases—to generate suspects and patterns of behavior. They plugged in details about the kidnapping and the system generated a possible suspect who had 53 prior arrests, several for child abductions. Police were dispatched to the vehicles associated with his name and the kidnapped child was recovered just half an hour after the abduction (Badger 2012).

One of the most valuable capabilities big data offers is real-time data collection. This is linked to real-time predictive analytics that enables real-time engagement of target audiences and real-time feedback. The capability for real-time data collection is already enabling new behaviors, even in its simplest forms. For instance, smart parking in large cities is beginning take hold, where parking sensors that collect data in real time communicate with other sensors to detect whether a car is in a parking space. The end user can interact with this data via an app to help expedite the parking process, lessening the amount of time spent driving around to find a parking spot and producing fewer emissions.

WHAT IS ANALYTICS?

Analytics is key to harnessing the power of big data. It is the process of connecting, analyzing, transforming, modeling, and visualizing data to discover valuable and actionable information. Jennifer Dutcher (2014) of the University of California, Berkeley's School of Information collected 43 definitions of big data from thought leaders across industries and found that central thoughts on big data are less about data and more about the tools used to parse and understand the data.

Two key activities take place during analytics. First, data are summarized and integrated to reduce their volume and converted into higher-level metrics and indicators that enable individuals to process the data. Second, the information generated is situated within the appropriate contexts in order to make sense of the data and conclusions to be rendered. Context, in this respect, means placing data within the appropriate domains from a functional expertise viewpoint (e.g., rules, historical data, future expectations) and being able to analyze these data by drawing on the necessary domain knowledge. This enables the creation of information and knowledge that drives decision making. As discussed in the next chapter, there are various analytical and modeling approaches that may be applied to datasets.

A critical issue that comes up when we discuss analytics of big data is the role of theory: Do we need theory or not? And, if so, when do we need it? A priori, or after looking at results for explanations? Much of the controversy around this topic started based on simplistic views in the early days of big data. Chris Anderson (2008) argued that due to big data there was no longer a need for theory (the title of the *Wired* article says it all: "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete"). According to Anderson, when there is enough data to describe a function of urban life, there then is no need for modeling or theory because the "numbers speak for themselves." Data, he argues, can now be computed to provide discoveries that science and hypothesizing cannot. Along this line of reasoning, correlation trumps causation without the needed explanation of a theory.

However, this argument has been challenged. Geoffrey West (2013) argues that big data does indeed need theory because without theory, data are meaningless and do not help us arrive at any new insights. When there are more data, analytics is likely to produce spurious correlations—and without theory to make sense of apparent relationships, incorrect assumptions can be made. Without theory, no one knows the critical questions to ask, which can spur unintended consequences and cause big data to lose its potency.

We must remember that big data can range from predictive to explanatory. The desire to see patterns that can help us predict the future does not trump our need to understand why the patterns being derived from big data are happening. Harvard professor Gary King asserts that data

ANALYTICS DRIVES FRAUD DETECTION

Fraud is an urgent problem that costs several industries billions of dollars each year. Credit card companies use real-time predictive analytics to detect fraud. When we make purchases, these companies use algorithms to detect anomalies based on previously known profiles of the user. The algorithms capture data points about our transactions to look for patterns in real time and quickly find the anomalies. Two things can happen when an anomaly is discovered. Predictive models using all of a customer's purchase data may classify a transaction as fraudulent and deny it on the spot. Or a manual review will take place and the customer will be contacted to verify the activity. Thus, if you usually spend $500 a month on your credit card to pay for gas, groceries, and bills in Columbus, Ohio, and then begin spending $3,300 a month on retail purchases in Dallas, the system will notice.

When using algorithms, there is always a chance of false positives. A false positive arises when algorithms recognize behaviors as fraudulent when they are not. This can be seen most commonly with credit cards when a card use is attempted, the card is denied, and the cardholder must contact the credit institution to alert them that the card has not been compromised. False positives can affect the development of smart cities due to the complexities of real-life situations—for instance, issues that can bring about false positives in pedestrian detection, such as shallow camera angles, poor lighting, and occlusions in fields of view (Silda 2009).

and analytics are not advanced enough to supplant traditional scientific methodologies and theories (Lazer et al. 2014). He notes that in analysis there needs to be confidence that the actual theoretical concept is being measured. This cannot be done without proper hypotheses and theory development beforehand. Thus, analytics coupled with theory is the correct approach to developing insights.

Discussions around theories can seem somewhat convoluted and complex, so this is what planners need to know about theorization in the planning process: A theory is a framework showing how a set of ideas is related. Theoretical knowledge on the topic is simply background information and perspectives. Thus, if a community is planning new bike lanes on city streets, knowledge of bicycle commuting trends, attributes of the physical environment that influence cycling, cycling safety measures on roads versus bike paths, motivations for cycling, and general and regional trends in urban mobility would be variables. Those variables should be investigated to uncover what has been studied about them, their dynamics, and how prior theoretical models might be valuable for consideration as analyses are conducted. The hypothesized relationships between variables are what emerge from the application of theory.

Obviously, theorization has not died; we still place an emphasis on theory. Simply mining data to find links and connections between data elements can be risky and can lead to the discovery of spurious relationships that could do quite a bit of harm to the design, implementation, and evaluation of policies and services. Many of the arguments against big data center on bad analysis and what some have called "big data hubris," which refers to data users' flawed analytics methodologies and the value given to flawed big data analytics outputs by users, which is likely to be transferred to others (see "Big Data Hubris: Google Flu Trends," p. 30).

Big data hubris centers on the notion that the value of big data has been overinflated to a point where individuals are finding value where there is none or are using big data as a substitute for traditional analytical methodologies. Big data hubris is not really about big data; it is more about problems with the analysis of big data. Many have wrongly believed that data and analytical tools are all that are needed to find insights; the truth is that more knowledge and expertise are needed to deal with big data.

In a traditional analysis, we would begin with theory to set up hypotheses. Hypotheses are educated guesses on how variables interact and the conditions under which these interactions take place. We would then use past experience and knowledge to justify our hypotheses and make a case for why they are novel. This is important because the next step requires data collection—which is not a costless effort—and then analysis. Analysis would lead to evidence in support (or in rejection) of our hypotheses.

One reason why we rely heavily on theory is to ensure that we are not randomly guessing about interactions and winging it. Another reason, and probably more relevant to this report, is that in the past, data collection required extensive effort. Surveys or interviews had to be conducted, data may have needed to be extracted and transformed from sources that required a lot of work (e.g., documents stored as images rather than easily editable PDF files), or the cost of recording data was high. In addition, analysis and computational resources were expensive and could not be used for testing out hunches and random assertions. Thus, theory is useful in designing a study and identifying factors to test within a given setting.

The prominence and placement of theory within the knowledge discovery cycle has undergone some revision. Today, we can simply mine data and discover associations between variables of interest. This is due to the fact that computational resources to do this are affordable and the requisite data are available. In addition, depending on the problem we are examining, we may not need to care as much about the intricacies of sampling as we may have access to data about the entire population. For example, if you wanted to study the online community engagement on civic platforms on Facebook, you may be able to get access to all group pages that are open, extract the data, and mine them. However, you should not make the mistake of generalizing from big data to populations that are not represented within the dataset, which can also happen even when full attention is paid to the attributes of the dataset. Knowing about and accounting for weaknesses in datasets is always important as well.

While we have the ability to run analyses cheaply, to make sense of the results we still need sound theoretical foundations and study designs. Otherwise, we will not be able to analyze and interpret the results in meaningful ways. Tyler Vigen (2015) is a student who has run various tests with data that showed high correlations but had no real connections. For instance, he found that US spending on space, technology, and science had a 99.7 percent correlation to suicides by hanging, strangulation, and suffocation. He also found that the divorce rate in Maine correlated to the per capita margarine consumption by 99.2 percent. David Leinweber found correlations between the S&P stock index and butter production in Bangladesh (Boyd and Crawford 2012). These examples of incorrect causation and correlation, while validated by the

METHODOLOGY PRIMER FOR PLANNERS

An appropriate method is extremely important to analytics. We recommend a mixed-method approach that utilizes qualitative and quantitative measures. "Data crunching" as described in this report is based on quantitative methodologies that focus on empirical measurements (e.g., surveys, meter reads, census data) and numeric-based analyses. In contrast, qualitative methods are based on data collected with greater richness in terms of the level of detail through such means as interviews, focus groups, or observations rather than analyses based on numerical outputs.

Qualitative methods are rooted in two types of reasoning approaches: inductive and deductive. Inductive reasoning makes broad generalizations based on specific observations while deductive reasoning uses hypotheses and theories to reach logical conclusions. When used together, quantitative and qualitative data can form more complete pictures (Creswell and Clark 2011). For instance, the Chicago Metropolitan Agency for Planning conducted a study on household travel and activities that included computer-assisted telephone interviews and GPS data collected from participants (Thakuriah, Tilahun, and Zellner 2015).

The mixing of qualitative and quantitative approaches helps to develop a greater understanding of the problems being researched. The messy and complex reality of cities means that if one method is used, there will inevitably be vital information missing. For instance, in the case of urban environmental sustainability, Frank Biermann (2007, 3) argues that computer modeling or visualizations will not tell the complete story of an issue, and he advocates for "qualitative, case-based, context-dependent, and reflexive" approaches in the development of research methods. Mixed methodologies also help prevent errors such as biases. These possible errors are discussed in more detail in Chapter 4.

BIG DATA HUBRIS: GOOGLE FLU TRENDS

The poster child for big data hubris is Google Flu Trends. In 2008, there was excitement about the ability of Google Flu Trends (Trends) to "nowcast" flu trends in the United States better than the Centers for Disease Control (CDC). Trends used a network analysis of online searches of flu symptoms to offer nearly real-time views of where people were getting sick; the central idea behind this was that when people are sick they tend to search for flu-related information through Google, which would provide indicators of overall flu prevalence.

At the time, Trends was thought to be doing much better work than the CDC, which was taking weeks to churn out information on cases because it was sending influenza samples to labs from primary care providers' patient encounters to find out where the flu was, what populations it was affecting, which strains were infecting people, whether antiviral drug resistance was occurring, and the like to create a more realistic picture of what was happening. However, there were critical problems with Trends.

The Trends methodology for assessing flu trends used 50 million search terms to find the best matches that fit into 1,152 data points. Therein lay the hubris: the methodology was problematic at the onset as it merged big and small data. Because of the large amount of data, the big data was overfitting the small number of cases. Trends staff were weeding out seasonal terms (in an ad hoc manner) that were unrelated but strongly correlated to the search term. This should have been an indication of an overfitting problem. The overfitting of unrelated terms that were bound to be correlated by chance and changes over time that modified search behavior were not properly managed (Lazer and Kennedy 2015).

As a result, over a four-year period, most weeks Trends overestimated flu trends. It overestimated trends by more than 50 percent from 2011 to 2012. Between August 2011 and September 2013, Trends overreported flu prevalence 100 out of 108 weeks (Lazer et al. 2014).

BIG DATA AT WORK

City governments are using analytics to capture various aspects of city life to enhance performance. In Dublin, Ireland, IBM has partnered with the Dublin City Council's Roads and Traffic department on a granular analysis of data to see when and how long buses are delayed at any location on a route in order to identify the root causes of traffic congestion. The city uses data from multiple intelligent traffic controls sources—such as inductive-loop traffic detectors, closed-circuit televisions, and onboard GPS systems—that receive updates from 1,000 city buses every 20 seconds. Equipping Ireland's buses with GPS units produces huge amounts of geospatial and activity data. Bus locations can be streamed in real time onto a digital map using geospatial and computing data. This visualization allows traffic controllers to view the current status at a glance. The controllers can then accelerate decision-making strategies when delays occur to prevent further congestion downstream.

In a city with high levels of bus ridership, finding ways to lessen traffic congestion for bus passengers is important. City planners found that Dublin's bus system was performing in a fragmented manner, with multiple buses running along the same routes at similar times and causing congestion. Through the use of granular analysis, variances in patterns were identified at different times of day; the details of these patterns then help planners make adjustments to the system, such as lane width, the operation of traffic signals, and driver behavior. The resulting interventions have led to about a 10- to 15-percent reduction in journey times (Tabbitt 2014).

In the pre-analytics era, city officials would have had to send employees physically out on bus routes for hours to understand bus trends, and those observations would still lack the depth that big data offers. With around 50 percent of people traveling in the Irish capital on public transit and 40 percent of those transit travelers taking the bus, these analytics are important to the mobility and quality of life of many people. The department is not only concerned with arrival times; it also wants to ensure that journey times are consistent and predictable, irrespective of the time of day. Several results from the pilot projects have been implemented, which has led to some service stability across Dublin within a two-year period.

statistical analyses, highlight the importance of the upfront placement of theory in traditional knowledge discovery (versus a post hoc analysis and interpretation of results).

With analytics, there can also be substantial unintended consequences. A significant benefit of big data is that there are large volumes of data that capture insights that are personal to us; that is what makes big data so valuable. However, those insights can be the result of discriminatory methods and practices, knowingly or unknowingly executed. Even with the best intentions, discriminatory algorithms can be put into place. It is easy to do.

For instance, following the 2008 recession, American Express began sending letters to some of their customers notifying them that their credit limit was being cut after analysis showed that the customers were shopping at stores frequented by people with poor repayment histories (Alloway 2015)—in essence, making customers guilty by association based on analytics. Eventually, American Express abandoned this practice, and the Credit Card Act of 2009 mandated further study of this type of activity. The use of analytics must be tempered with strong ethics and standards about what lines are not to be crossed.

Figure 2.1. The concept of the Internet of Things (Chesky_W, Thinkstock)

THE INTERNET OF THINGS: OPPORTUNITIES FOR MORE DATA

Through smart data and Internet of Things (IoT) technologies, more growth will happen in areas involving variables that previously were difficult to measure. IoT is the interconnection of devices that "talk" to each other, including personal electronics, sensors, and networks. These devices will monitor specific conditions, such as motion, vibration, and temperature. The promise of IoT technology is a revolutionized, fully connected world where the environment and people are connected by objects (Figure 2.1).

By 2020, 50 billion IoT physical objects with sensing and communicating power are set to come online (Best 2016). Technology companies are more innovative than ever and are finding ways to make users' lives easier through the embedding of sensors for monitoring. Cities are benefiting in a variety of ways, such as increasing efficiency. In San Antonio, Texas, for example, city traffic lights were connected through IoT technology to assist with traffic management. Cities are also saving money—money that can be directed toward other projects. In Barcelona, Spain, the city saves $58 million annually by using smart-water technology (Maddox 2015). IoT innovations are also helping cities become much more intuitive. Anticipation is an important method of reducing risk and making better decisions. In Melbourne, Australia, Yarra Trams uses IoT technology to update its train fleet. The operations center and maintenance teams anticipate the travel needs of residents and tourists by turning elements of the tram system into data points, with sensors that report in real time about equipment, service, and maintenance disruptions and issues (Roberts 2013).

For urban planners, IoT opens up a whole new realm of possibilities. Using this information, city infrastructures can be reimagined, which will have impacts on social, political, and environmental policies. Connected cars will make driving safer and cut down on carbon emissions while public spaces can be adapted and adjusted to users' needs to create entertainment, educational opportunities, and interactive spaces for gathering, collaboration, fellowship, and further innovation. This will all be done through insights gleaned from the collection of big data and analytics, insights that will be used to connect all facets of daily life. This might seem far-fetched now, but imagine 20 years ago, when floppy disks were used to save computer work and when people used cameras with actual film that had to be developed in order to share physically printed photos with other people.

Connectivity between machines and devices is changing the world. IoT technologies are the fastest growing of Internet-

DISPARATE IMPACT

Suresh Venkatasubramanian of the University of Utah's School of Computing conducts research to determine if software algorithms can exhibit bias that is consistent with the legal classification of disparate impact. Disparate impact is the idea that a policy is discriminatory if it has an adverse impact on any group based on race, gender, sexual orientation, religion, or any other protected status. He does this analysis through algorithms (ironically) that test if they can accurately predict a person's race or gender, even when hidden. If they can, there is a significant chance of the potential for bias and a basis for disparate impact. In instances where he finds bias, he notes that a solution is to simply redistribute the data in a way that will prevent the algorithm from seeing the pieces of data that are used to create bias (University of Utah 2015).

A related example is the Supreme Court case *Texas Department of Housing and Community Affairs v. Inclusive Communities Project, Inc.*, which addressed disparate impact. The federal government provides tax credits to home builders that build low-income housing, and the credits are administered through a designated state agency. The Inclusive Communities Project sued the Texas Department of Housing and Community Affairs (the agency designated to administer tax credits for Texas), arguing that the agency was administering the credits disproportionately to African American inner-city areas and to too few white suburban areas. To support its claim, the Inclusive Communities Project used statistical analysis to prove that the credit program favored these inner-city communities (92.29 percent of credits went to housing projects in these areas), which had the result of actually segregating Texans by race (Kirchner 2015). In a five-to-four ruling, the Supreme Court justices found that disparate impact claims were valid under the Fair Housing Act.

SMART-CITY INNOVATION AROUND THE WORLD

Glasgow, Scotland

The new and innovative ways to harness all of the data cities have access to mean that there are new opportunities to solve significant community challenges. The City of Glasgow in Scotland uses big data and analytics to drive smart-city solutions. Glasgow is the United Kingdom's most "deprived" city, with the lowest life expectancy of all UK cities (Arnett 2014). Characterized as the "sickest city" in the United Kingdom, Glasgow has significant challenges with healthy lifestyles (Ash 2014; Reid 2011). In a 2011 comparative study of Liverpool and Manchester (two cities with equally low employment levels and high deprivation and inequality averages), Glasgow residents were 30 percent more likely to die young, with 60 percent of deaths attributed to drugs, alcohol, suicide, and violence (Glasgow Centre for Population Health 2016).

Despite this, the city has made a concerted effort to become a smart city by spending £24 million on smart technology. The city has installed intelligent streetlights that brighten and dim depending on activity levels, record air pollution and weather details, and operate closed-circuit cameras (Macdonell 2015). The roads are equipped with sensors that read and analyze data quickly to modify traffic patterns and prevent bottlenecks. The city is even becoming socially smart: city maps are even equipped to show, for example, where people can find recovery and support groups and services, cultural organizations and events, and cycling routes and infrastructure (http://map.glasgow.gov.uk).

Additionally, Glasgow is focusing more on prediction and intervention to drive its smart city. Much of the prediction revolves around crime; for instance, researchers found that men were more likely to perpetrate domestic violence on the nights that the city's football team plays. The city hopes that identifying such patterns and predicting behaviors will allow for interventions before problems occur.

Finally, Glasgow's open data program has been a key feature in its efforts to become a smart city; the city's information is located on an intelligent-operations platform that stores and analyzes data in real time. City officials are releasing data on measures such as the number of deaths registered in Glasgow, the number of first-time mothers, and rates of alcoholism in specific areas, in the hopes of finding unexpected relationships between social and environmental issues. Anyone can go on the website (https://data.glasgow.gov.uk/dataset) and download and analyze the data, or use the online dashboards, apps, visualizations, and widgets.

Copenhagen, Denmark

In Copenhagen, Denmark, ambitious goals of environmental sustainability and technological connectedness have propelled the city into being one of the best smart cities in the world. Copenhagen has made great strides already in meeting its goals by having completed the following accomplishments:

- Installed a new district cooling system where cold is extracted from harbor water, which saves 70 percent of energy compared to traditional air conditioning
- Cleaned the harbor, making it attractive enough for businesses that harbor activities actually generate revenues; the harbor is now so clean that citizens can swim in it
- Integrated transportation and cycling solutions to reduce congestion and improve health
- Invested in public bikes in 2005 that has resulted in 45 percent of citizens biking to work or school daily
- Developed an efficient district heating system with 98 percent of all households connected
- Achieved recognition within a highly digitalized country ranking second to the Netherlands when it comes to broadband penetration in the European Union

One of Copenhagen's most ambitious goals is to become the world's first carbon-neutral capital by 2025 (Copenhagen 2016).

THE FUTURE OF THE INTERNET OF THINGS

What are the predicted developments around the Internet of Things (IoT) and data generation capacity?

- In 2015, over 50 percent of IoT activity was related to manufacturing, transportation, smart cities, and consumer applications. In the next five years, this activity will be vertically diversified, with all industries rolling out IoT initiatives (IDC 2014).
- In 2018, over 25 percent of all government external spending on IoT will be by local governments (Clarke et al. 2014).
- The amount of data created by IoT devices will increase from 134.5 zettabytes per year in 2014 to 507.5 zettabytes per year in 2019 (Cisco 2015).
- IoT data are underutilized; it is estimated that just 1 percent of data being generated from IoT is being analyzed (Manyika et al. 2015).
- IoT will have 26 billion units, or objects, installed by 2020; the revenue for IoT product and service suppliers is expected to exceed $300 billion (Gartner 2014).
- Within five years, 40 percent of wearable technology will evolve and become viable consumer mass market alternatives to smart phones (Clarke et al. 2014).

connected sensors. They are projected to have an annual $11 trillion impact by eliminating the need to manually do rudimentary tasks like household chores because smart appliances will manage those undertakings and free up people's time to do other things (Manyika et al. 2015). IoT's growth can be attributed to the fact that "things" can be any objects (living or inanimate) to which sensors can be attached or embedded in, and most of these data are transmitted machine to machine. Also, IoT is the wave of the future because of low-power sensor technology at low cost, widespread wireless connectivity, and large amounts of storage and computing power at affordable prices. Cement is an example of a product that has not seen many technological revolutions until recently, with the advent of "smart cement" and the ability to transmit valuable information. Smart cement is being designed to use highly sensitive materials that sense changes in the environment, such as contamination, stresses, cracks, fluid loss, temperatures, and pressures, all in real time. In the future, this technology will allow the cement to "talk" to planners and engineers and help minimize building errors and accidents like potential building and bridge collapses.

Wearable Technology

Wearables, or wearable technology, is an IoT technology made up of clothing and accessories with embedded sensors and software that can connect to other objects or devices without human intervention. New technology gadgets at CES (Consumer Electronics Show) 2016 revealed IoT products reflecting trends that are expected to grow, such as the Owlet smart baby monitor, a washable, wearable baby monitor that tracks babies' oxygen levels and heart rates and displays them on smartphones, or Garmin's Varia Vision device for cyclists that mounts onto sunglasses and provides information in the user's line of sight, including turn-by-turn directions and even phone notifications.

You might think of older, less-sophisticated wearables such as sleep apnea test machines, heart monitors, or calculator watches, but today's wearables are much more advanced because they collect vast amounts of data and can be used in new and innovative ways: as activity trackers (e.g., Fitbit, Jawbone), media devices (e.g., Apple Watch, Moto 360), or devices that provide communication between the wearer and other devices (e.g., Google Glass). These devices are documenting users' activities, collecting data, providing information, and helping users achieve predetermined goals or make decisions.

Traditionally, data collected from wearables are for users only. They are often collecting additional, detailed information on users, including GPS logs and data on heart rates, activities, and routines. This kind of information is fine for many users, and actually are data that are fun to have. For instance, a Reddit user posted to a thread about an issue with his wife's Fitbit noting that it was reflecting a higher resting heartbeat than usual, reaching up to 110 beats per minutes, and was logging 10 hours in the fat burning zone, which he felt was impossible based on her activity level (Dier 2016). He attributed it to a sensor problem with the Fitbit. Another Reddit user responded to his thread and asked if his wife was pregnant. Not knowing that a higher heartbeat was related to pregnancy, the Reddit user and his wife took a pregnancy test and found out they were expecting.

However, the collection of such personal data can be used against users by cybercriminals or insurers, or even in legal proceedings. A study conducted by Accenture found that 63 percent of insurance executives believe that wearables will be adopted broadly by the insurance industry in the next two years, which could be used for underwriting, marketing, risk management, product development, workers' compensation, and personal auto injury claims management purposes (Insurance Journal 2015). Usage-based insurance could be a new trend where auto insurance premium costs are dependent on vehicle use and usage is monitored through a device on vehicles. According to Accenture (2014), most insurers have already begun commercial pilot programs to test approaches to usage-based insurance. Further, cybercriminals can use wearable data to attack users. A 2015 study by Symantec found that 100 percent of wearable devices were trackable and transmitted user-generated data (e.g., names, passwords, email addresses) without encryption (Symantec 2015).

GOING FORWARD

Undoubtedly, data are an asset. However, value can only be captured from data through analytics. Right now, smartphones are the generators of most data, but this will soon change, and they will be one of many sources of data for planners. IoT technologies will revolutionize the way in which cities connect infrastructure, the Internet, and people. In order to realize the promise of big data and analytics, planners should envision the transformation of their cities and find the smartest ways to implement the changes.

As urban planners help lead their communities into the future, visioning exercises can help cities with the development of missions and goals (for example, "in the next 13 years, the city will be a leader in the Southwest of smart-

DRONES

Drones are one of the fastest growing Internet of Things technologies in terms of sales, use, and popularity. NASA used drones for high-altitude research as early as the 1970s. Back then, drones were a novelty and very expensive, which kept them out of the reach of many would-be users. Today, cheaper and more advanced drones—which can have high-quality sensors, GPS, and autopilot features—have made them much more accessible (Figure 2.2). Drones now offer a range of new and interesting capabilities, such as assisting farmers with inspecting crops from new vantage points and with more frequency, delivering packages, and contributing to search and rescue missions.

Drones are a critical way to collect data that ordinarily would take months and many rounds of data collection to gather. Urban planners are increasingly using drones to understand landscapes and the condition of infrastructure and to monitor environmental changes in real time. In Saarbrücken, Germany, for example, building inspectors are using drones to determine building conditions. The drones have sensors and advanced imaging and recording capabilities that create 3-D data models through which inspectors and engineers can identify building defects or cracks. In the past, inspectors would need cranes and scaffolding to inspect the variety of buildings and hard-to-reach areas. Once defects were noted, they would have to note the trouble spots on a 2-D map, which was an error-prone process. The use of inspection drones helps to reduce the time of inspections by only taking a few hours to complete what would ordinarily take days. One 15-minute drone flight can yield nearly 1,200 photos that are then combined to create 2-D and 3-D models (Global Construction Review 2014). However, there are limitations to this process—a key one being that the drone has to be operated manually and remain within sight of the controller, which limits how high it can go.

Even with these limitations, this only scratches the surface of what people are doing with drones. The spatial data that drones collect provide accurate, up-to-date information that can be used in various ways. Drones are good for gathering high-resolution aerial images over small areas. In most cases, spatial data collected from drones can be merged with other datasets to facilitate modeling, mapping, planning, environmental monitoring, disaster risk mitigation, and infrastructure development. Thus, drones are quickly reducing the barriers to mapping tasks that have usually been done by specialized firms, allowing them to now be done in house.

Further, data from drones are also being shared in novel ways. Spatial data are being crowdsourced to increase usage and efficiency. For example, the National Park Service initiated a crowdsourcing campaign to find missing hikers at Craters of the Moon National Monument and Preserve in Idaho. Due to time constraints and rugged terrain at the park, missing hikers are hard to spot. The National Park Service began posting aerial shots from drones online for people to look through to see if they spotted hikers who could be in distress (Robinson 2013). Drones were also used after the massive earthquake in Nepal in 2012 to help assess damage and map and document the states of cities, and they have been used in numerous other search and rescue efforts (Selkirk 2016).

Even though drone technologies are useful, they have many legal hurdles to overcome as they become more accessible. Issues surrounding drones include air safety, regulation, privacy, the use of drones as weapons, and drone insurance. Most states have introduced some type of drone legislation but little has been passed. In general, US policies and regulations related to drones have been contested. More work is needed to ensure appropriate regulation of this technology.

Figure 2.2. Hovering drone (agnormark, Thinkstock)

city adoption"). The visioning process should include multiple stakeholders, both public and private, to ensure that realistic plans are being made. We contend that a city with aspirations of becoming a data-driven smart city cannot do so without visioning the future with all necessary parties.

After visioning has taken place, it is time to implement. Implementation requires clear-cut strategies for achieving the goals envisioned for the city. Without this, efforts can be uncoordinated and expensive, and they run the risk of working against the city's vision. There are many issues that can curtail the implementation of the city's vision, such as lack of funds, low interest, incompatible goals amongst stakeholders, criticism, and disasters. To power through this, the visioning at the onset must include the proper stakeholders, and be feasible and well thought out. From there, strategies of achieving these goals should be identified for each section of the vision.

A LITTLE CREATIVITY

Today, the most popular wearable that many of us use are fitness trackers. Fitness trackers are a personal technology that helps users understand their movements and reach goals. But they have other uses, including uses for planners. Planners seeking to enhance walkability and decrease use of automobiles understand that the walking patterns of people in a community provide important information for future planning. Paul Goddin (2014) of Mobility Lab used a wearable fitness tracker in four cities to conduct a personal experiment about walkability that provided interesting results with urban planning implications. In New York, he walked 19,000 daily steps; in Washington, DC, 12,000 steps; in Baltimore, Maryland, 7,000 steps; and in Poquoson, Virginia, 4,000 steps. The walkability of New York and Washington, DC, has a lot to do with physical access to many different amenities, including public transportation, as well as with the cost of owning a car. Conversely, he notes that Poquoson is a typical American suburb lacking public transportation and the pedestrian access that allow people do such activities as run basic errands without driving. As Goddin's project demonstrates, the use of wearables could assist planners in finding ways to increase walkability and improve urban sustainability.

Wearables are not infallible and their accuracy has been the subject of lawsuits and speculation. Rachel Metz (2015) explained the variation in accuracy between her fitness trackers (Apple Watch, Microsoft Band, and Polar H7 Bluetooth chest strap) on her bike ride to and from work. For heart-rate measurements, each device at times measured nearly the same heartbeat per minute; at other times there was a 77 heartbeat-per-minute difference. For calories burned, figures were inconsistent as well. Her experiment showed that these devices were not strictly accurate but rather depended on characteristics of the wearer (e.g., hairiness, sweatiness, blood flow to wrist, even the presence of tattoos). For planners, these inconsistencies could aid in figuring out how to use wearables as part of their arsenal of sensors to gather data on citizens' behaviors.

AUTONOMOUS VEHICLES

The Internet of Things technology that is most likely to change in the next 20 years is autonomous vehicles (AVs). AVs are driverless vehicles that will be equipped with GPS technology, high-definition cameras, infrared and radar scanning, advanced control and sensory systems, and algorithms that use data from this equipment to operate a vehicle—all without human intervention. Driving the development of AV technology are its potential benefits, including the following:

- Avoiding traffic collisions that are usually caused by human error
- Reducing traffic congestion
- Providing vehicle-to-vehicle and vehicle-to-infrastructure communica-tion
- Reducing the need for public safety employees as safety regulations and laws are complied with more consistently
- Reducing the need for physical parking spaces because AVs will run in constant car-sharing loops
- Providing faster travel times

A few years back, AVs were considered just an idea—a hope of technology enthusiasts. Today, it is abundantly clear that AVs will be a game changer in the near future. In 2016, the US National Highway Transportation Safety Administration (NHTSA) determined in a landmark declaration that Google's self-driving car artificial intelligence is considered an actual "driver" (Urmson 2015). NHTSA has noted before this recognition that AVs bring with them "completely new possibilities for improving highway safety, increasing environmental benefits, expanding mobility, and creating new economic opportunities for jobs and investment" (NHTSA n.d.). Clearly, we are fast approaching a new mobility paradigm that has major implications for planners everywhere.

One of the first impacts of AVs will be on the transportation infrastructure. Modifications will be needed to enable communication between vehicles and urban infrastructure, particularly to increase safety. AVs will also change urban life in terms of car ownership. These vehicles will operate themselves, and this, together with emerging trends around carsharing, will likely mean that fewer people will need or want to own vehicles.

AVs will also change the functioning of different sectors. For example, AVs are projected to be far safer than human-operated vehicles—which means that there is likely to be a dip in government revenue from moving violations and parking tickets. Insurance-related issues, such as liability, will need to be reassessed. The use of AVs will also present environmental changes if the cars are made more ecofriendly (through the use of electricity instead of gasoline) and less environmental damage is incurred. More information on this technology is available in *Local Government 2035: Strategic Trends and Implications of New Technologies* (Desouza et al. 2015).

CHAPTER 3

AN OVERVIEW OF ANALYTICAL APPROACHES

Critical to the ability to leverage big data are analytical techniques. Traditional analytical techniques have to scale to handle the volume of data involved when dealing with big data. In addition, analytical techniques must be able to integrate different types of data and operate across a wide assortment of systems and platforms. This chapter reviews some of the most commonly used analytical approaches for big data.

TAPPING INTO OPEN-SOURCE TOOLS

The cost of analytics can range from free to expensive. For those planners not looking to break the bank on a new set of computational tools, there are open-source tools that are available at no cost. Open-source tools have source code that is openly published for use or modification from the original design—free of charge. For instance, GRASS (Geographic Resources Analysis Support System) GIS is a free, open-source GIS software suite that provides image processing, spatial modeling, visualization, and data management software (https://grass.osgeo.org). It is a project of the Open Source Geospatial Foundation, an organization supporting collaborative development of open-source geospatial software. GRASS GIS is currently used by academics; government agencies, such as the National Aeronautics and Space Administration, the National Oceanic and Atmospheric Administration, and the US Census Bureau; and private-sector environmental companies.

DATA MINING

Data mining, as the name implies, helps us discover latent patterns and associations between variables in large datasets. There are a multitude of analytical techniques that can be applied to uncover information from databases, including association rules, decision trees, and classification models.

From a process point of view, generic data mining works the following way: the dataset is split into two (the training set and the validation set). First, the analyst constructs a model that detects patterns or associations in using the training set. The analyst will normally explore several models depending on the nature of the dataset. Models will then be refined to improve predictive accuracy. Second, the model is then applied to the validation set to gauge its predictive accuracy and performance. Models that are deemed good have good predictive capacity and also have explanatory power (i.e., how much variance in the dependent variable is explained by the variables in the model). Association rules (or association mining) is a data mining technique that is probably the most well-known and straightforward. With this technique, one can discover the presence of an element in a dataset as it relates to the co-occurrence of other elements. For example, if a person buys milk and bread, it is highly likely that the person will also purchase eggs.

A 2016 report from the Data Center, a resource center in southeast Louisiana, revealed correlations between low-income housing vouchers and gun violence based on datasets using American Community Survey data and US Department of Housing and Urban Development Housing Choice Vouchers Program and subsidized housing data. The analysis found that children living in homes secured through housing vouchers are in highly racially segregated neighborhoods with heavy exposure to violence. In fact, of the nearly 19,000 children in the housing voucher program, 55 percent lived in the 12 voucher-assisted neighborhoods that had experienced an average of 10 shootings annually, while only 2 percent lived in the 13 neighborhoods with zero shootings (Greater New Orleans Fair Housing Action Center 2016).

In another example of data mining, researchers used transit data from the Canadian Transit Authority to mine user behavior data based on smart card usage (Agard, Morency, and Trépanier 2007). The smart cards collected a variety of data from users as they traveled: (1) the rider's card is validated when a passenger boards a bus, (2) the system categorizes the passenger based on the bus's planned route, (3) the GPS reader identifies where the passenger boarded, (4) the system validates that the route is at the correct location, and (5) the system collects other identification data (including card number, time of boarding, validation status of the passenger's card, and stop number). The data mining exercise resulted in the identification of four clusters of users with similar travel patterns. The largest user group, almost half of travelers, was made up of people making regular trips during peak travel hours (79.4 percent in the morning and 71 percent in the evening). The percentage of riders traveling during off-peak times was very low (6.4 percent during the day and 2.6 percent in the evening).

Data mining has also been used to enhance road safety. Vision Zero is a multinational road safety policy designed to promote smart roadway behaviors and designs that anticipate mistakes and ultimately prevent severe injury or death. This approach has been adopted in a number of US cities, including New York; Portland, Oregon; Chicago; San Francisco; and Los Angeles. The City of Los Angeles has used a data-driven approach to attain its Vision Zero goals. Using injury data, the city developed the High Injury Network (HIN) to identify streets and intersections with high concentrations of traffic collisions that resulted in injuries or deaths (Figure 3.1 and Figure 3.2). The HIN revealed that over 65 percent of all fatal and severe injuries involving pedestrians happen on just 6 percent of city streets. (The HIN visualizations and dataset along with other Vision Zero maps are available at http://visionzerosf.org/maps-data.)

MACHINE LEARNING

Analytical approaches use algorithms to iteratively learn from data and build (and update) models. Machine learning approaches normally take one of two learning approaches: supervised or unsupervised. In supervised learning, as the name implies, the model is developed by analyzing patterns using a set of inputs and outputs. The analytical tool continues to learn and the prediction accuracy of the model improves over time. For example, an analysis might look at a set of inputs, such as credit history, time of a transaction, transaction amount, and location of the event, and an output, such as whether fraud was committed. A model is developed for various combinations of the inputs and their interactions and their abilities to correctly predict if a transaction is fraudulent or not. Once

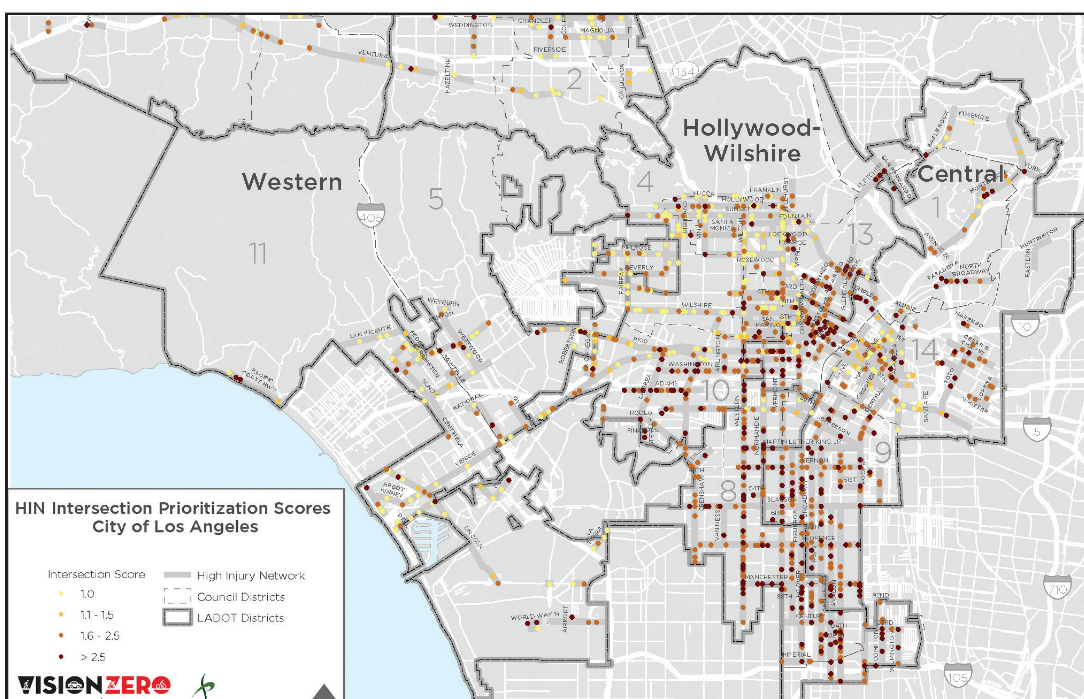


Figure 3.1. High Injury Network intersections in Los Angeles (Vision Zero Los Angeles)

the model is developed and put into operation, the system uses the opportunity to continuously learn every time the model is deployed to improve the prediction accuracy. In unsupervised learning, data elements are not labeled and we allow the algorithm to identify patterns among the various elements. The machine normally will have to identify the latent structure in the dataset through a clustering approach to identify groups and clusters.

Just as with data mining, several analytical techniques can be used for both supervised and unsupervised learning depending on the size, structure, and format of the dataset. Machine learning techniques are built to process large amounts of data; these great quantities of data are necessary to generate sufficient levels of accuracy. By ingesting large amounts of data, the algorithms can continue to refine the precision of outputs as they consider a greater set of connections between the various data elements and as their knowledge of the search (problem) space continues to increase. Machine learning approaches are now being incorporated into next-generation systems called cognitive computing systems.

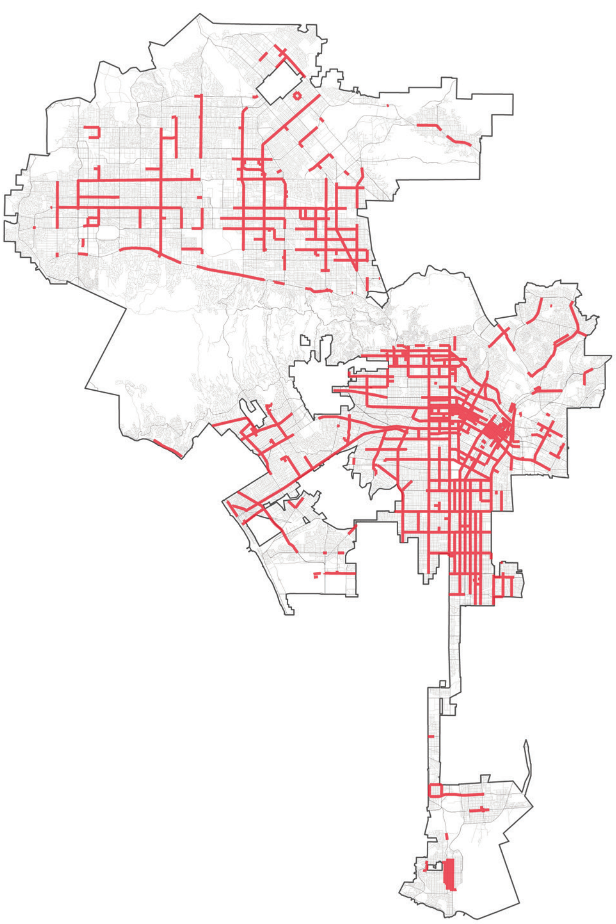


Figure 3.2. Streets with a high concentration of traffic collisions that result in severe injury or death (Vision Zero Los Angeles)

The applications of machine learning to analyses of big data are varied, with significant implications for policy and planning. One example of machine learning involves "selfies" and citizen involvement helping to advance scientific knowledge. In Singapore, citizens are acting as scientists by uploading their outdoor selfies to a new smartphone app project called AirTick. Graduate student Pan Xheng Xiang designed AirTick to collect photos from users that help assess the haziness of the environment—the direction, duration, and intensity of sunlight—and check this against official air quality data. Through machine learning, the app will eventually be able to predict pollution levels based solely on the selfies. The potential of this technology will be increasingly important since the number of selfies is on the rise (Rutkin 2016); a 2015 industry study indicated that the average millennial will take about 25,700 selfies in a lifetime (Glum 2015). The AirTick project is significant because it could serve as an alternative to expensive sensors and is another opportunity for citizen science.

Fire rescue and crime predictions are examples of a data-driven approach where machine learning algorithms can help make communities safer. The fire and rescue department in New South Wales, Australia, is incorporating machine learning techniques into its work with a new system called Miinder (Pennington 2014). Miinder predicts when an emergency will most likely occur by using geographic and variable data such as weather forecast and crime data and will send a crew out to proactively manage an emergency that is likely to happen. The system will also go back and analyze whether that crew produced the most optimal outcome and will make adjustments to the ruling system for future incidents. The Miinder-based calculations are designed to improve response and cut costs by increasing administrative efficiencies such as linking systems to dispatch the closest truck to an emergency rather than one stationed in the vicinity. The adoption of the Miinder system has involved upgrades to the technical equipment of the department's 400 administrative staff, including laptops, mobile phones, and other portable devices.

The Japanese technology giant Hitachi has developed a system called the Hitachi Visualization Predictive Crime

Analytics, which processes large amounts of data and hundreds of variables (e.g., social media conversations, 911 data, license plate readers, gunshot sensors, weather reports, public transit maps) using a machine learning system to find patterns that are presented geospatially and that are oblivious to humans (Hitachi 2015). The locations can be pinpointed down to 200-square-meter spots, and the system assigns a level of threat to places and situations where crime is likely to occur. Hitachi's system has not been deployed yet, but there are plans to have trial runs in several unspecified cities to assess the accuracy of the system and whether it unfairly targets and profiles innocent people instead of criminals.

SENTIMENT ANALYSIS

Today, given the popularity of social media and the proliferation of commentary on these platforms, sentiment analysis (or opinion mining) has become a popular analytical approach. At its core, sentiment analysis leverages techniques in natural language processing, computational linguistics, and text analysis to uncover information from a large body of text.

Machine learning approaches that focus on sentiment analysis are often used when assessing polarity. Opinions, gathered from sources such as tweets and Facebook posts, are placed in a huge bucket of words that human coders have decided they would like to study. The algorithms not only work to classify words on various dimensions (e.g., positive/negative, level of authenticity, confidence) but can also be used to uncover latent connections between words and collections of words (i.e., phrases).

The most common application of this approach is to uncover the sentiments associated with data on social networks. Does the piece of text represent positive or negative attitudes? Sentiment analysis focuses on identifying what attitudes people hold towards a given organization or entity, object (e.g., a product), or agent (e.g., a politician). This type of analysis can reveal how attitudes differ among sets of users and possible links between various attitudes. In addition, given the nature of social networks, analysts are also interested in learning about the transformation and persistence of attitudes and sentiments over time and what might be done to modify them. Sentiment analysis, however, is an extremely subjective method of classification and the quality of analysis can vary heavily based on the nature of data, the level of care taken to clean and structure the data, and the level of knowledge that the analyst has in the subject matter.

Sentiment analysis can also be used to study the confidence of sentiments. For example, if people share negative or positive opinions on a given issue, how confident can we be about their opinions? Generally speaking, sentiment analysis algorithms assign positive or negative polarity scores to words, sentences, or blocks of text. Lexicons are developed that categorize words according to the polarity (i.e., words are classified as having positive, negative, or neutral sentiments). Additionally, lexicons identify the strength of the polarity, often on a -10 to +10 scale.

More advanced models of sentiment analysis can help decision makers parse large quantities of text data to classify them into various categories—for example, if examination of data on election campaigning conversations shows that a given piece of text talks about issues related to healthcare, terrorism, or economic opportunity. Once texts are classified into various categories, analysts can take a deeper look at the intricacies of features such as sentiments, polarity, and framing. As these tools increase in sophistication, urban planners will find them incorporated into all sorts of systems: examples include mobile apps that provide data on how to navigate urban spaces, traditional web-based citizen engagement platforms, and even self-service technologies for delivery of public services.

Sentiment analysis is likely to be a tool that will be used for an innovative plan to connect the federal government and citizens. In 2015 Yelp, the popular web and mobile service that allows people to rate local businesses, announced that the US government will have pages on the site that it will review (Rein 2015). In an effort to become more transparent and responsive to taxpayers, the new Public Services and Government section will be used by agencies to interact with the public and also to bolster the Obama Administration's digital strategy. Eventually, data will be used to drive improvements to government services.

Another example of sentiment analysis involves city social media strategies. Researchers conducted sentiment analyses of 125 cities with active Twitter accounts. They placed each city's Twitter handle into the Mississippi State University Social Science Research Center's Social Media Analysis and Tracking System (Zavattaro, French, and Mohanty 2015). Tweets to and from each of the cities were tracked for five weeks. In total, there were approximately 17,222 tweets, with 4,779 tweets being separated out because they were tweets from the local governments or advertisements. After analysis, the researchers found that 41 percent of tweets were positive, 49 percent were neutral, and 10 percent were negative. They found that most city governments were largely using their Twitter accounts to push information (including meeting announcements, traffic alerts, office closings, and cultural events) to the public and that they only maintained one-way relationships with citizens. They found more positive sentiment in cities that encouraged active participation with citizens by posing questions, retweeting, and sharing photos.

GEOGRAPHIC INFORMATION SYSTEMS

A critical element of big data analysis is the layering of data elements to get richer contextual information on environments. From a spatial perspective, leveraging geographic information is critical. Geographic information systems (GIS) deliver various forms of information as multilayered maps in which each layer provides different information. For example, LIDAR (light detection and ranging) is a remote sensing technology that combines light pulses (from pulsed lasers that measure variable distances) and other data to produce stereo imagery in the form of high-resolution optical remote-sensing images. This technique is being widely used to classify and model cities in 3-D. Objects can be ranked using LIDAR, and building parameters—such as roof type, height, and number of houses—can be defined. Additionally, LIDAR allows users to calculate a number of different parameters related to structure, including average building sizes, the floor-space index, vegetation fraction, the percentage of impervious surfaces, and building density. This technique also offers the possibility of mapping urban biotopes.

Following 2012's Hurricane Sandy, the largest Atlantic hurricane to hit the East Coast, LIDAR technology was used by the US Geological Survey (USGS) to document changes to affected areas (Sopkin et al. 2014). This documentation was critical to the protection of infrastructure and to public safety. Extreme storms, such as hurricanes and winter storms, can generate dangerous waves that create dune erosion, overwash, and inundation of water that can reshape the coastline. The changes in the coastline can then destroy buildings and infrastructure and even result in injury or fatalities. LIDAR systems mapped hundreds of miles of coastline in one day. Analysts continued this process in different areas numerous times after the hurricane. Through comparison of pre- and post-storm images, USGS was able to assess that overwash and erosion occurred in several places—some heavily populated and developed. With this information, officials were able to make informed decisions on appropriate remediation of the damage caused by the storm.

Another example illustrates how GIS can engage citizens in the planning process. Low-income neighborhoods in Detroit are benefiting from big data and analytics to help mitigate housing stress in the city. Loveland Technologies created a $1.5 million program called Motor City Mapping (www.motorcitymapping.org) to tackle a challenge from the Obama Administration: move or clear all blighted structures and lots as part of the Detroit Blight Removal Task Force initiative. In order to begin, data managers and urban planners collected data on all buildings and vacant lots in the 142 square miles of the city through property surveys and by using a smartphone app. The app uploaded information on the project using photos and texts, or "blexting" (blight + texting), to a database with over 375,000 parcels of land. The Motor City Mapping project added these data to a usable online interactive map for citizens and stakeholders to visualize the city at the neighborhood and street levels. Analytics focused on a "vacancy index" that combined multiple datasets, including water and sewage use data and DTE Energy and US Postal Service data. Now these data are available to the entire region, and this dataset allows for substantive analysis that can help people make informed decisions rather than basing them on assumptions and hypothetical thinking.

NETWORK ANALYSIS

A critical aspect of big data analysis is to discover connections between elements. The exploration of connections helps us understand the structure and evolution of networks. Social network analysis has been around for several decades and looks specifically at the connections between people. It can also be used to understand connections between groups (e.g., teams) and organizations (e.g., businesses, countries) as well. A form of network analysis is link analysis, which looks at connections between objects.

A common measure in network analysis is the degree of centrality, used to determine the relative influence of each user on the network as a whole. Prominent users are identified by a technique called node categorization (or sizing nodes). Nodes are categorized based on the degree of centrality, which is reflected in the number of edges that are connected to the node in consideration. Modularity is a metric to measure connectivity between clusters within the network. A highly modular network will have dense

ESRI: GEOGRAPHIC INFORMATION SYSTEMS APPLICATIONS

Geographic information systems (GIS) have evolved from complex and difficult applications to essential parts of how citizens and governments understand and relate to their environments. Increasingly, public agencies are using GIS to capture, store, analyze, and manipulate data to decrease costs, increase communication, and improve decision making. The advancements in visualization and information technologies have further accelerated the use of GIS in the public sector in fields such as emergency management and urban planning.

Esri (Environmental Systems Research Institute) (www.esri.com), an international GIS company founded in 1969, is a leader in helping governments across the world use location-based data for improving decision making and engage citizens in the policy process. In 2015 the company has 350,000 clients, including governments, businesses, and nonprofit organizations. Esri is changing the way governments conduct their operations and connect with citizens. It is revolutionizing the use of data in developing innovative solutions for traditional governance challenges in areas such as traffic control, emergency response, and health care.

During Hurricane Ike in 2008, the emergency response team across the tri-state area of greater Cincinnati realized that its paper-based maps were ineffective in assessing the scale of damages for response preparedness. The Ohio-Kentucky-Indiana Regional Council of Governments, Hamilton County Emergency Management Agency, and the Cincinnati Fire Department partnered with Esri to create the Regional Asset Verification and Emergency Network (RAVEN911) (http://raven911.net). RAVEN911 layers information about critical infrastructure systems (e.g., fire stations, hospitals, schools, power grids) with real-time data (e.g., social media feeds, traffic flows, weather reports) to handle emergency responses. For instance, during a chemical leak, the team can enter information such as the location of the leak, type of leak, and type of device to calculate the impact of the leak on nearby residential areas and to establish evacuation areas. Response teams can use RAVEN911 through computers or mobile devices.

In 2014 the City of Portsmouth in Virginia conducted a community health survey to understand the health status of vulnerable populations (Kurkjian et al. 2016). It used the Community Assessment for Public Health Emergency Response technique to collect baseline information about communities in order to determine which areas needed interventions. It collected data by separating the city into 35 clusters, or housing units, using census blocks and picking seven households at random from each cluster. In-person surveys were conducted with the randomly selected households and collected using ArcGIS. This allowed survey responses to be recorded and marked by location for post-survey geospatial data analysis. The resulting baseline data provided policy makers with an understanding of the communities' health status before a disaster had occurred.

In addition, Esri also makes data accessible through ArcGIS Open Data (www.opendata.arcgis.com), a subscription-based mapping platform. Open data increases government transparency and also provides opportunities for innovation and increased citizen participation. With these data, communities can generate branded open data websites with visualizations that are searchable and downloadable. Users can also combine open data with other datasets in ArcGIS Online (www.arcgis.com) or add their own data to do advanced geospatial analysis. For example, the US Environmental Protection Agency's MyEnvironment website (www3.epa.gov/enviro/myenviro) is hosted by ArcGIS Online and provides the public with access to data from federal, state, local, and private data sources on environmental indicators such as air and water quality, energy, climate, and health (Esri 2011).

THE USE OF SPATIAL EFFICIENCY METRICS TO IMPROVE INFRASTRUCTURE SYSTEMS PLANNING AND PERFORMANCE EVALUATION

Tony H. Grubesic, Director, Center for Spatial Reasoning & Policy Analytics, Arizona State University

Infrastructure systems, broadly defined, include all of the basic organizational structures and physical facilities required for the day-to-day operation of society. The physical facilities include transportation infrastructure (e.g., roads, rails, airports), the electrical grid, gas and water pipelines, and myriad key assets such as fire stations, dams, and emergency warning sirens. The challenges associated with expanding, contracting, or improving the efficiency of infrastructure systems can be daunting, especially when the agencies responsible for building, operating, and maintaining these systems are budget constrained.

A powerful but underutilized family of measures for infrastructure system planning and performance evaluation relate to spatial efficiency. Spatial efficiency refers to the process of assessing costs associated with a given locational arrangement of a service and comparing them to costs associated with the best-known alternative arrangement (Desai and Storbeck 1990; Fisher and Rushton 1979). Identified differences can provide guidance for improving the existing locational arrangement, as well as establishing benchmarks for system performance under a given arrangement. More importantly, because the growth or contraction of a region is not perfectly predictable, efficient location decisions at one point in time may not reflect efficiency at another time. As a result, when the costs of providing basic infrastructure services inflate (e.g., students are bussed further) or the quality of these services degrade (e.g., fire station response times increase), the relative efficiency of the system can be tracked over space and time and strategies for improving location decisions can be implemented accordingly.

A recent study explored the concept of spatial efficiency in the context of rural air transportation (Grubesic et al. 2016). Essential Air Service (EAS) is a federal program that provides subsidies to commercial carriers to provide air service to rural and remote US communities. Community eligibility is determined by two factors. First, if the local airport had commercial service prior to deregulation in 1978, it is eligible for EAS. Second, the local airport must also be more than 70 miles from an existing large or medium airport, as defined by the Federal Aviation Administration. EAS is controversial because it is expensive—costing taxpayers more than $200 million per year. It is also underutilized, with many planes flying without any passengers.

This study evaluated the spatial efficiency of the existing EAS airport configuration, its degree of redundant service, and the marginal benefits associated with reduced levels of service to identify the best-known alternative arrangement of EAS communities. Specifically, by maximizing the potential demand (population) covered by the EAS network (spatial efficiency) as well as maximizing the relative operational efficiency of the system by evaluating passenger load factors (capacity utilization and subsidy allocations), a measure of combined efficiency was established that enabled tradeoffs between spatial and operational efficiencies and benchmarks for performance. One of the unique features of this application was the use of geographic information systems (GIS) to manage, manipulate, generate, and visualize spatial and operational information for the mathematical programming models. For example, GIS was used to determine the 70-mile network catchment areas around each airport and to capture characteristics of the underlying population (using census block groups).

The EAS airport network in the continental United States served over 30 million people during 2011 with a budget of $216 million and an average operational efficiency of 76 percent. Using a simple set of budget reductions (e.g., 10 percent, 20 percent), a subset of EAS airports were defunded, but this was done in a way to maximize spatial efficiency, operational efficiency, or a balance of both. For example, a 10 percent reduction in the EAS budget reduced total funding to $195 million. When emphasizing spatial efficiency, 12 of the 116 EAS airports were defunded, leaving 104 airports operating. These 104 airports were able to serve 99.7 percent of the population served by the original EAS configuration, saving taxpayers over $21 million dollars. When larger cuts were made to the system, such as a 40 percent budget reduction, 79 airports remained funded, reducing the taxpayer burden by more than $86 million. Surprisingly, when emphasizing spatial efficiency in this scenario, a full 94.49 percent of the potential demand for EAS service remained covered.

In sum, infrastructure planning and performance evaluation is a complex task involving a wide array of constituencies, decision makers, and stakeholders. Evaluating the spatial efficiency of infrastructure systems and their performance is becoming more important as the costs associated with building, operating, and maintaining these systems increase.

Figure 3.3. Number of jobs reached in a 10-minute walkshed from each building in Cambridge, Massachusetts (Image: City Form Lab)

connections within a cluster and sparse connections between nodes of different clusters.

Network analysis allows us to see the connections among various data elements. For example, we can see the connections between two agents (humans) or two objects (devices) as well as those that occur between humans and objects (e.g., two humans who are connected to or dependent on the same object—say, an energy source). Today, urban planners must become comfortable dealing with large datasets where computation of networks is not as straightforward as when only a single form of interaction or connection was considered. One of the attributes of big data is the ability to link datasets (i.e., connect data elements) across various domains.

The City Form Lab at MIT has created a state-of-the-art ArcGIS toolbox for urban spatial network analysis called Urban Network Analysis (UNA) to help urban planners and geographers with the spatial configuration of a city (City Form Lab 2016). UNA finds the relationship between people or places along spatial networks by assessing distances, encounters, and accessibilities in a city through geometry, topology, and building features (e.g., population, significance of the building to the larger urban landscape). The centrality tool finds associations in areas such as reach, gravity, betweenness, closeness, and straightness. The redundancy tool calculates the redundancy index, redundant paths, and a wayfinding index. UNA is available for ArcGIS 10 and Rhino 3D modeling software. The UNA toolbox offers powerful methods. It is useful for small-scale detailed network analyses of dense urban areas and can be used to easily scale for sparser large-scale regional networks (Figure 3.3).

AGENT-BASED MODELING

Imagine being able to simulate traffic patterns by simulating the actions of each driver. This is possible through agent-based modeling (ABM). ABM is a computational method that simulates an automated agent's behaviors and analyzes macro-level patterns derived from micro-level agents (Zheng et al. 2013). These are computational models where complex systems are developed by specifying the behavior of agents (individuals or organizations) and their resources, behaviors, preferences, and interactions as well as activity with other agents, objects (artifacts such as buildings, systems, and land), and the environment. Each agent

CONSTRUCTING COMPUTATIONAL MODELS USING AGENT-BASED MODELING TO UNDERSTAND POLICY RESISTANCE

Spiro Maroulis, Assistant Professor, School of Public Affairs, Arizona State University

Actions intended to improve economic, social, and organizational problems often provoke surprising and unanticipated responses by the system a particular management or public policy is trying to improve. Examples of such "policy resistance" abound: additional highway lanes have often increased congestion in part by stimulating new demand for the highway; widespread use of antibiotics has led to more dangerous pathogens by precipitating the evolution of drug-resistant strains; California's mandating of smaller class sizes led to the sudden hiring of unqualified teachers as the supply of teachers struggled to keep up with the spike in demand (Sterman 2000).

Why is policy resistance so prevalent? The short answer is that the preceding examples are all instances of actions taken within complex systems where collective outcomes are often an aggregation of highly interdependent individual actions, effects are often disproportional to cause, and cause and effect are often separated in time and in space. Consequently, learning from the school of hard knocks is rather slow and difficult, and mental models of these systems are often incomplete.

Constructing computational models that capture the underlying structure of complex systems can accelerate learning and improve related mental models. Agent-based modeling is a particularly useful approach for modeling social and economic systems that represents the structure of the system as a set of heterogeneous agents, rules that govern the behavior of those agents, and the topology of their relationships with each other or their environment (Wilensky and Rand 2015). By constructing and running worlds of interconnected agents over a series of discrete time steps, one can discover the emergence of macro-level properties from the individual-level actions of those agents as well as use the simulations to identify leverage points in a social system—points where small, local changes can have disproportionate system-level impacts. Agent-based models have been used to better understand a wide variety of policy-relevant topics including urban land use (Huang et al. 2013), environmental conservation (An et al. 2005), and educational policy (Maroulis et al. 2010).

For example, Maroulis et al. (2014) use this approach to better understand education reforms that attempt to introduce competition into school systems by giving families the ability to choose schools other than their assigned neighborhood school. The agents in the model are high schools and students, sampled from longitudinal data on Chicago public schools, placed on a grid of sites that represent the geography of the city (Figure 3.4). Initially, all students attend their assigned neighborhood high school. In subsequent time periods, incoming cohorts of students are allowed to start choosing using a preference function based on the mean student achievement and geographic proximity of a potential school. Over time, some schools gain students while others lose them. Schools whose enrollments fall below a minimum threshold of enrollment are permanently closed. At the end of a time period, new schools open in random geographic locations within the city.

Analysis of the model revealed the importance of considering the dynamics, and not just equilibrium characteristics, of the transition to school choice. More specifically, it demonstrated that the equilibrium outcome of transitioning to choice depends not only on the quality and total number of new schools in the system but also on the timing of the entry of the new schools.

Good intentions often produce unanticipated consequences. However, we do not need to resign ourselves to the fact that social and economic systems are full of surprises. Computational models can help us discover and anticipate those surprises by exploring scenarios that build upon, but go beyond, the data at hand.

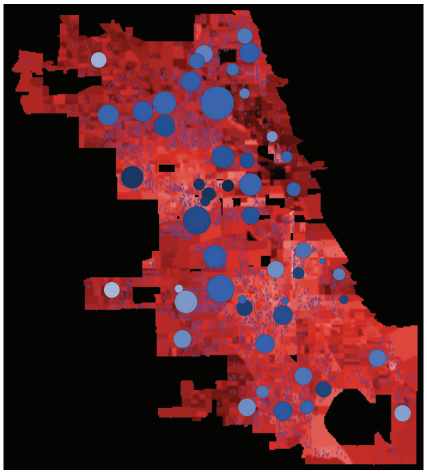


Figure 3.4. Map showing schools, proportional enrollment, and mean achievement (Adapted from Maroulis et al. 2014)

corresponds to a real-world actor who is cognitively and socially bound to a place. The simulation allows the tracing of collective patterns emerging from the interactions between agents and their environments; the interaction rules, constraints, and environmental conditions are modeled to represent the options being considered.

The interactions between the individuals and objects are the key to the simulation. When building ABMs, it is important to keep model specifications simple and clearly specify the agents, their resources, their activities, the rules of interaction between agents, and influences from the environment. This prevents the development of overly complex models that are hard to understand. ABMs also give modelers the potential to turn on and off and change the intensity of various variables to see how outcomes might change. For example, the density of a city can be changed from low to high or the presence of a particular incentive from on to off.

In the setup of an ABM, organizational parameters or environmental features are established. Then the simulation is run, which leads to the generation of large-scale data that need to be analyzed. This allows for the testing of multiple policy interventions and their intended and unintended consequences. Simulation allows system dynamics to emerge. For instance, modeling urban land use with ABM can lead to better land-management decisions because it provides a better understanding of issues like urbanization, re-urbanization, urban sprawl, and urban shrinkage and how these outcomes are shaped by decisions at the individual level (e.g., decisions by residents about where to buy homes) and at the policy level (e.g., economic incentives offered by local governments to builders and real estate investors).

The benefit of ABM is the ability to model what cannot be easily tested in the real world due to the cost and capacity necessary to do so. For example, we can build models to understand how smallpox might spread through a city and the effects of various inoculation strategies to contain the spread. Here ABM is helpful, as trying to test this out in the physical world is not possible. Similarly, we might want to simulate how an economy would react to a policy intervention, where, even if we could test it out in the real world, the cost of getting it wrong is too high. ABM also allows for significant flexibility along multiple dimensions as more agents can be added, the complexity of agents—including the ability to learn and evolve, the degree of rationality, behavior, and rules of interaction—can be modified, and the levels of description and aggregation can be adjusted.

At the same time, there are challenges to using ABM. If the models are not established accurately, there will be significant issues with the simulations and the resultant interpretations. Also, knowns and unknowns always exist in establishing the models because of the complexity and irrationality of human behavior, subjective choices, and complex psychological processes. In addition, to power simulations (especially at the city level), planners need good-quality data in order to produce realistic models with trustworthy results. For instance, the data needed to model city roads could include travel data from electronic transit cards and rideshare apps, accident data, mobile phone density information, and retail transaction data. The inclusion of these data means that the simulation need not rely on assumptions of the population but rather is a true representation of what is happening.

Agent-based modeling has been used for urban design. To understand the influences of neighborhood street layouts, for example, researchers applied ABM to a set of typical layouts spanning the last 100 years of neighborhood design in Ottawa, Ontario (Jin 2010):

- **Traditional grid:** Represents historic cities in the era when walking, carts, streetcars, and horses were common means of transportation
- **Postwar suburban:** Presumes the existence of (and accommodates) private cars
- **Traditional neighborhood:** Sheds uniformity by avoiding four-way intersections
- **Fused grid:** Reflects a conventional suburban grid with a hierarchy and limited four-way intersections

Each layout has the same experimental settings in terms of characteristics such as density and population. Each layout is also the same size (half mile by half mile) with local facilities and exits located in identical spots. Density is medium level at 16.2 households per neighborhood. The study showed that neo-traditional and fused layouts were more pedestrian friendly. Shorter walking distances to facilities, less traffic, fewer crossings, and more social interactions characterized these neighborhood designs. This research demonstrated that ABM can provide a useful link between design, implementation, and outcome.

GAMING

Planners can also use big data to capture activity, interest, and interaction through gaming, a useful and fun method of experimenting and decision making using gamification,

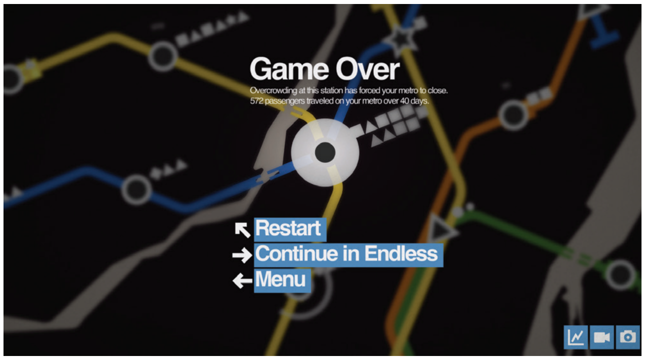


Figure 3.5. Mini Metro interface (Dinosaur Polo Club)

or the use of game elements in nongame contexts. Popular games used for urban planning are SimCity and Mini Metro. SimCity (www.simcity.com) is an open-ended game series that tasks players with developing a city—using features such as zones, types of buildings, sewage systems, and placement of public safety units—while maintaining a stable budget and sustaining resident happiness. In the transportation game Mini Metro (http://dinopoloclub.com/minimetro), players must develop an efficient subway system in a growing city (Figure 3.5).

The point of gaming is to always understand the behaviors of users and what motivates these behaviors. Gaming is a benefit to urban planning because it takes user activity, captured through big data, to understand interactions and decisions. Data collected from games can aggregate such issues as common problems that arise, people's most immediate decisions, what motivates behaviors (e.g., instant gratification, incentives), and what was learned on subsequent tries of the game. Further, gaming can also help transform big data into smart data. This happens through turning user decisions into data-driven insights.

As planning-based games become popularized, there are distinct data opportunities hidden in the behaviors of players. In traditional games such as Farmville, every move the player makes is logged and tracked by the provider. The provider is then able to understand features of the game that were challenging for the player, places where the player stopped playing (possibly indicating frustration or boredom), or exciting parts. This essentially becomes a feedback loop between the player and the developer and can help inform planning by identifying common mistakes players made, the solutions they developed, and other moves made during the game. Other in-game metrics can be collected that further assist developers and planners in understanding players' actions and how these relate to planning activities and strategies.

The success of gaming all depends on the quality of the game (e.g., instant feedback, instant results), how it attracts and sustains user interest, and the timeliness of insights drawn from user actions (e.g., interaction levels, time needed to complete challenges, connections between public data such as tweets). Using gaming is a serious and time-consuming endeavor because of the details needed for games to be interesting to users. One prediction suggests that 80 percent of gaming solutions will fail to attain their objectives due to poor design (Gartner 2012). For quality designs, planners will likely need to join forces with experts in this area to find gaming solutions that are affordable and attractive to users and can provide actionable insights.

Gaming has been applied to planning through civic gamification, a method of using gaming to tackle urban issues through educating, informing, and allowing real-time decision making by citizens. This form of urban engagement is traditionally done on open platforms where users can connect with one other to enhance the process and outcomes. Detroit 24/7 is a game that allows citizen to think, learn, and share their ideas about Detroit's future while competing to earn points and prizes. Detroit Works Project Long Term Planning collected ideas from citizens and used the insights to guide the development of a 50-year strategic plan for the city (Block 2012).

BUILDING INFORMATION MODELING

Building information modeling (BIM) is a process of digitally representing the physical and functional characteristics of places such as buildings, tunnels, wastewater and electrical facilities, ports, prisons, and warehouses. BIM is a departure from the 2-D technical drawings that are used in traditional building design. It extends beyond 3-D to a fourth dimension of time and a fifth dimension of cost. BIM provides opportunities to use the Internet of Things (IoT) and other tools, such as GIS to combine and analyze physical and administrative data (such as vacancies and lease space) with other data sources such as LIDAR information.

BIM is a better solution for larger projects because of the increased level of information found in the models, increased confidence in the work (especially with data from IoT), and decreased amount of time needed to work on the models. Unlike other modeling programs, a change made to one aspect of a model will update the whole system, which limits the time needed to work and rework different aspects of the model. BIM also conveys consistent and coordinated representations of digital models. As different views or contexts are toggled, the information remains reliable; this is expected to increase as more smart-city and IoT data are infused into these systems. As IoT technologies become more pervasive throughout government and city infrastructures, they are expected to help with advanced technologies like BIM by augmenting knowledge, increasing productivity, sharpening user skills, and increasing precision in the field.

In Arizona, an interdisciplinary team dedicated to saving the Grandstand Building at the Arizona State Fairgrounds (a building that dates back to the New Deal) from being demolished banded together to use BIM and other tools to preserve it. A key way to save the building was to create an accurate as-built Historic American Building Survey for fundraising and preservation efforts; the survey will also go into the US Library of Congress. With no money and little time, the team used drones, 3-D laser scanning, and point clouds in ARCHICAD 19 (topographical information) along with BIM to create billions of data points and develop a survey in a matter of days. It was estimated that traditional surveying methods would have involved 15 to 20 surveys and three months to complete at an estimated cost of $250,000 (GRAPHISOFT 2015).

CONCLUSION

This chapter described tools that planners can use to analyze big data. It also described simulation, modeling, and gaming approaches for understanding complex systems and testing interventions—from large datasets that need to be analyzed to arrive at actionable insights. While all of these will not be feasible or valuable for every organization, one or two alone or a series used together could help planners achieve their goals in smarter ways. For instance, in London, BIM, GIS technologies, laser scanning, augmented reality tools, and digital data capture were deployed to revolutionize the way rail projects are conceptualized and developed (Crossrail 2016).

To reap the benefits of any of these tools, each one must be carefully evaluated to ensure that it is used in the right context and that it is able to deliver the desired outcomes. As planners familiarize themselves with these tools, they should assess whether they are appropriate for their organizations. In addition, the many online courses and instructional videos on each of the topics mentioned in this chapter are greatlow-cost or no-cost ways to get started.

ARE PLANNING DEPARTMENTS PLANNING FOR BIG DATA?

Shannon McElvaney, Esri

I attended a talk at a geodesign summit a while back titled "Where Are the Planners?" It turned out that the talk was really about sensors for parking stalls. The point was that in all of the speaker's presentations about smart parking there were never any planners in the audience. I found the question intriguing, and it got me thinking about smart devices and big data and their role in planning.

I am now convinced that big data and the Internet of Things (IoT) will inevitably change the way we do planning. Think about it. Everything you have to plan for—water, disaster response, traffic, jobs, housing—can be improved upon with big data. We have already seen the rise of big data in the form of smart-city initiatives, with an increasing number of sensors and systems collecting greater and greater amounts of data. This started first in the public works, transportation, and emergency management sectors, but the stream of big data is beginning to trickle into planning departments as well.

The management of water is one area driving the growth of IoT. Companies like Valarm (www.valarm.net) are creating web-based dashboards connected to field devices that track remote tank or well levels and water usage from flowmeters. These data help water resource managers know how much water is available at any given moment. This takes the guesswork out of supply and demand questions. It also creates a great historical data trail for trend and alternative scenario comparisons, which are essential to planning.

Understanding threats is another important aspect of planning. Valarm is applying the same technology to monitor fire risk by assessing fuel moisture, precipitation, relative humidity, solar radiation, barometric pressure, temperature, wind speed, and wind direction (Figure 3.6) (Pultar 2016). When these data are coupled with location, topography, and historical sensor information, the result is a greatly improved early warning system that helps firefighters, foresters, and natural resource managers better manage environments and protect life and property.

Live data tied to location gives us a better understanding of how the environment works, and this can help planners, the public, and insurers improve changes to land use and zoning codes and help ensure that the intended outcomes are realized. Say a developer wants to build in a single-family residential area near a wildland interface prone to fire and flooding. The developer wants the municipality to build the infrastructure to support this development and offers a few million dollars to cover some of the buildout costs, including some flood mitigation. Is this a good deal for that county?

If the municipality had valuable historical condition information for the area along with data about its geology, rainfall, and vegetation cover, it could assess the true costs and benefits of that infrastructure development. Property assessment, tax revenue, and maintenance over 20 years, along with flood risk, could easily show whether that municipality's investment would be enough to cover the true costs of infrastructure over the development's design lifetime (McElvaney 2012).

With extreme weather becoming the new norm, planners are recognizing with greater frequency that it may be more effective and less costly to locate

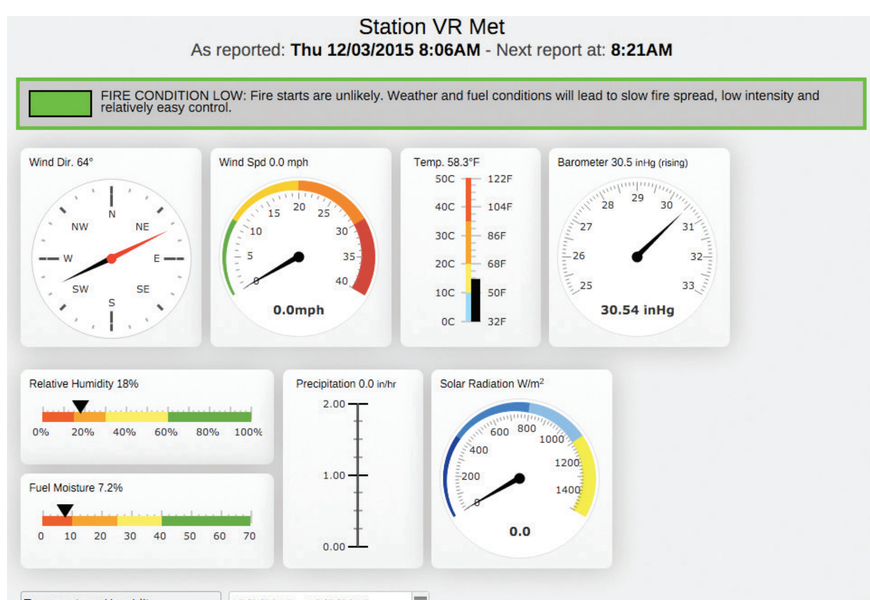


Figure 3.6. Cloud-based dashboard for a sensor that remotely monitors fire risk weather in Southern California (Valarm, LLC)

new developments outside of hazard areas, rather than attempt to control the hazard itself through mitigation. Modern technology trends—like the use of environmental sensors developed by Valarm and other companies—will give planners and decision makers the scientific data they need to help developers and the public understand the true costs of development in marginal areas. This helps everyone make smarter decisions.

Many agencies are starting to put live data feeds online. A good example is traffic feeds that let users pull live traffic data into their maps. Another example is live feeds of bus routes showing the locations of buses and their arrival times at specific stops. Analyzing actual traffic data using location and temporal components can also reveal patterns that are extremely useful to transportation planners. That is how contraflow lanes and the timing of traffic lights to ease congestion came about. And that is just the tip of the iceberg.

What if you want to design a vibrant neighborhood? It helps to quantify how people use and move through public spaces. Benchmarking before and after measures helps confirm planning assumptions with hard data, so planners can justify needed street or sidewalk treatments, or even parklets or alleyway improvements, to increase both walkability and pedestrian traffic. Sensors can help planners with this. New technology companies like Placemeter (www.placemeter.com) build video sensors that track and measure the volume of pedestrian and vehicle traffic, walking directions, store visits, and even the wait time at sites like food trucks. This is big data at work in local places—driving both active living and economic development.

Parking is another area of planning where we are seeing an explosion of applications and tools. Sensors in the pavement are connected to phone apps that make it easier for people to find and book a parking spot, even before they arrive at the spot. The parking meter even alerts the driver if it is about to expire and asks whether the driver would like to buy more time. Technology applications such as this pose questions to planners. For example, how should apps like this affect the parking requirements for new developments?

Then there are always the benefits to be gained by enriching big data with additional demographic data to gain even greater insight. Never has there been such a variety of data available to identify trends and understand who is affected most by a particular development or zoning change. As cities open up and share more and more of the data they already capture, we will see an explosion of innovation to help us create better, healthier, and more sustainable communities.

Taking the big idea of innovation hubs that connect people, data, and tools from the regional to local levels is exemplified by Esri's Smart Communities effort and, in particular, the Esri Hub Program (www.esri.com/smart-communities). The City of Los Angeles is an early adopter that has unlocked its geodata with the Los Angeles GeoHub (http://geohub.lacity.org), which puts data on a common platform and then shares it with city departments, neighboring municipalities, and the public. In this way, the big data already collected by the city can be used by others and "mashed" together with other data to increase government effectiveness while also improving residents' quality of life. The platform is also a way to inspire entrepreneurs to start their own companies that design and build helpful applications using these open data sources.

The City of Boston's Department of Neighborhood Development is combining Esri location data and spatial analytics with its SAP HANA operational data to improve city services across departments, including the planning department (SAP 2016). Crime data and 911 calls are tied to parcel data to help identify problem parcels related to blight. Business analytics and operation dashboards are exposing inefficacies and have helped the city decrease its permitting backlog from 600 overdue permits to 3. At each step of the way, analytics are saving the city time and money in the office so staff can devote more time to building the community.

Big data is made more powerful when it is location enabled, open, and shared across departments and with the public. It exposes synergies and cost savings in unexpected places. And data-driven decisions help take the guesswork out of planning for a desired outcome. How do you see planning departments taking advantage of the rise of big data?

CHAPTER 4

A FRAMEWORK FOR LEVERAGING BIG DATA THROUGH ANALYTICS

As detailed in the previous chapter, big data offers us the opportunity to gain new insights and perspectives on urban issues. It also affords us the opportunity to innovate how we design, plan, and govern our communities. The analytical process needed to use big data requires rigor and attention to detail. This process needs to ensure that data are collected from the most appropriate sources, validated, and integrated. In order to do this, appropriate analytical techniques need to be employed, analytical outputs should be carefully scrutinized, and relevant insights and actions should be implemented and disseminated in an optimal manner.

Bloomberg Business Week Research Services conducted a survey in 2013 of 103 top managers (C-level executives and business executives) in government agencies and found that 81 percent of managers "strongly agreed" that big data was crucial to achieving their missions. However, these managers' agreement did not translate into action (Mullich 2013). A chief reason for this is that several obstacles prevent quick big data technology uptake; these issues need to be addressed before big data is used. Obstacles include resistant organizational cultures, lack of expertise within organizations on how to conduct analytics, and a lack of resources to push the "data agenda" forward.

The overall process and following discussion can be broken down into four phases. Phase 1 focuses on managing the sources of data. Phase 2 examines the intricacies involved in running analytics. Phase 3 looks at issues one has to consider when making sense of analytical outcomes. Finally, Phase 4 considers how to communicate and implement outcomes that result from analytics and how to build the capacity for ongoing renewal and refinement of your analytical activities. This framework is adapted from a knowledge management framework outlined by Desouza (2006).

PHASE 1: MANAGEMENT OF DATA SOURCES

Data are generated by agents (i.e., humans) and objects (i.e., things). Traditionally, planners were confined to a limited number of sources from which they could collect data. These sources produced data on a regular basis and in defined forms, and they were deemed credible due to their official designations, such as the US Census Bureau, or the regulated processes they followed. Today, things are quite different. In addition to traditional sources, planners have available data from a much wider variety of sources. Many of these sources do not have the same credibility or authority as traditional data sources. They also often lack permanency and emit or produce data that are of varied types, frequencies, and formats. Source management is about knowing what sources one should pay attention to, being able to evaluate the credibility and veracity of sources, extracting data from these sources when needed, organizing sources, and protecting sources and the data being extracted (Desouza 2006).

Know Your Sources

Knowing your sources and their characteristics is an important activity for planners using big data. Broadly speaking, sources can be grouped into two categories: internal (those within the planning organization) and external (those outside the planning organization). Sources can also be grouped by the domains they represent (e.g., transportation, utilities, health and well-being). While these details may seem trivial, most planning departments struggle with creating a list of sources and prioritizing these sources based on the nature of the data available. Planners will want to do an exhaustive environmental scan to unearth valuable data sources. For each source, capture details such as what kind of access you have to it; the nature of data emitted by the source, including details such as frequency and type of data; and the level of sensitivity of the source, which will determine how to secure the data.

It is also important to remember that managing sources is not a costless function. Dedicated resources will be needed to ensure that you are able to retrieve, analyze, and protect data from the source. As such, if you choose the wrong source or an inferior source, you are paying an opportunity cost (i.e., you might be giving up the opportunity to consider other more valuable inputs as you have already made commitments to one source).

Once you know your sources, you must do the painful task of mapping them to the problems they will help you solve. Today, we have become lazy when it comes to our discipline in collecting data. Given the drastic decline in the cost of data collection and abundance of storage, we collect more data than ever before. Much of the data is never organized, analyzed, or archived appropriately. However, as noted previously, while the cost of collecting and storing data is low, it is not zero. Storing and securing data, from a technical and organizational resource perspective, is not cheap. You need to have a goal in mind—a problem you want to solve or an opportunity that you want to realize—when the data are collected. A random strategy that involves collecting anything available is not helpful for several reasons. First it is costly, as noted previously. Second, it leads to information overload and paralysis in terms of analytics. Third, you will not be able to identify the added value each source of data provides you when it comes to making a decision.

Organizing Your Sources

Organizing sources calls for categorizing and linking sources, which helps with the visualization of information space. Linking sources allows you to triangulate data from multiple human and machine sensors. It also provides you the capacity to achieve greater awareness of what is happening through data fusion. For example, the City of New York's Office of Policy and Strategic Planning (the "geek squad") combines data from different departments to solve pressing urban problems. It combined catering certificate data from the New York Business Integrity Commission with geospatial restaurant location data to develop a list of potential restaurants to track for illegal oil dumping. Using this list, the health department was 95 percent successful in tracking illegal dumpers (Feuer 2013).

Mapping information space is vital to understanding gaps in sources—where there should be internal and external sources, but there are none. Knowing the gaps helps with decisions about how to go about acquiring new sources. For instance, the city of Rio de Janeiro in Brazil partnered with the mobile application developer Waze to collect real-time transit data from citizens. The Waze travel app collects a wide range of information about users, such as trip origin, time, date, speed, destination, and route (Figure 4.1). Further, users through the app can report road conditions and accidents. Rio de Janeiro is combining this information with its traffic data to manage traffic flows, repair roads, and prepare accident responses.

When purchasing data, it is critical that you study the data sharing agreements carefully. Data sharing agreements

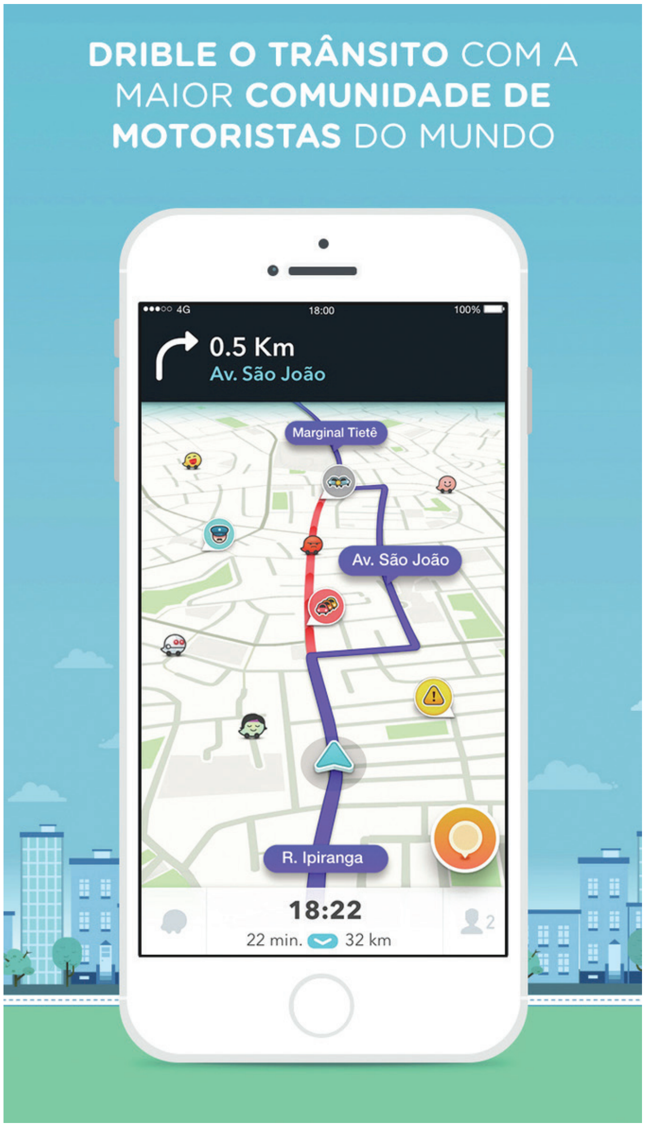


Figure 4.1. Waze app from Brazil (Waze)

SENSOR DATA

Sensor data comes from objects embedded within our environments, and given the advancements in technologies, they can be embedded within our physical environments with ease. These sensors can capture basic details, such as when a person is in a room, but can also go a step further and capture the movement of objects or the environmental conditions around an object (e.g., temperature).

As we develop smart cities, designers and planners will exploit data from sensors to get real-time situational awareness of these environments. This awareness provides us with an input that allows for adjustments based on changes in the environment. For example, smart tolling based on data emitted from sensors in cars on roads is already commonplace in many cities. Depending on the time of day and, in some cases, the amount of traffic on the road, the toll prices vary. In an even more advanced application, sensors in cars are being developed to monitor driving behavior by analysis of head and eye movements and even to predict what you will do next (Knight 2015).

Sensors can be embedded in objects (e.g., physical infrastructure) or on agents (e.g., humans, animals,) and can emit data of various types (e.g., text, numeric, video, images). The greater the distance between a sensor and the physical environment, the lower the level of detail that the sensor can provide. For example, a satellite rotating in orbit can provide detail about our physical environment on Earth, but it is not able to gather data at the level of granularity that can be extracted from an intelligent sensor embedded within an object.

Currently, the most important sensors are mobile devices (e.g., phones, intelligent wrist bands). People leave digital traces of themselves when they use mobile phones. These data provide a window into the lives of residents and allow urban planners to understand trends such as population distribution, activities in the city, connections between different areas, and commuting patterns. Call detail records (CDRs) that include information about inbound and outbound calls—such as originator, terminator, start time, length of the call, type of call (call or texting), and the route of the call—were traditionally used for only operational purposes such as billing.

Today, the same data are being used more creatively. For instance, the data can help us understand movements of users; cities around the world are partnering with mobile data providers to understand trends, particularly transportation and mobility trends. In the Ivory Coast in West Africa, transportation managers are using a model developed by IBM to reduce residents' travel time based on CDRs. A cell phone emits a use and location signal, which is registered with a cell phone tower that reports the user's general location. Movement is then captured as mobile data is transferred to new towers that signify new locations. Using these data, planners in the Ivory Coast deployed new modes of transportation (e.g., large buses, mini-buses, shared taxies) that better met the travel and mobility needs of residents (Talbot 2013).

We also now have sensors embedded in objects (e.g., cars) in ways that are shifting paradigms. These sensors leverage wireless technologies to transmit data between other objects. For instance, in 2014 the National Highway Traffic Safety Administration announced new efforts to enable vehicle-to-vehicle communication technology in vehicles, a move representing the next generation of auto safety (NHTSA 2014). Vehicle-to-vehicle communication allows the use of networks to transmit information between vehicles so they can "talk" to one another and exchange safety information—such as basic safety data, speed, and position—at a rate of ten times per second. With this type of data, both drivers would be able to understand what is happening sooner and avoid collisions and congestion as much as possible, and transportation planners would be able to more quickly and effectively manage the transportation network.

As we move toward urban transport environments with autonomous vehicles, the significance of vehicle-to-vehicle and vehicle-to-infrastructure sensors will only increase. As with all new technologies, there are opportunities for and threats to users. In 2015, researchers Charlie Miller and Chris Valasek remotely hacked a jeep that was in motion on the road to prove that it could be done (Greenberg 2015). They used a hacking method called a zero-day exploit, which targeted Jeep Cherokees, to acquire wireless control of the vehicle and command of the vehicle's steering, brakes, transmission, and entertainment system dashboard. The remote capabilities were all available at the controllers' fingertips. Later in this chapter, the issue of security will be discussed.

Humans as Sensors

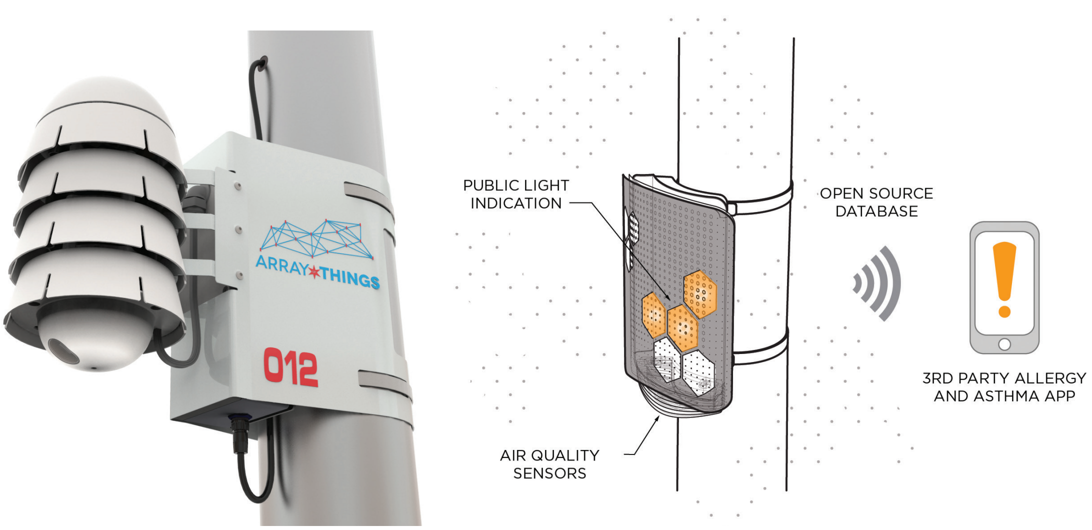
A vast data source is data collected using humans as sensors. Humans create a wide array of data either passively or actively. It is important for planners to appreciate the nuances of the various kinds of data. First, there are data that humans actively create and share on purpose. These data can be transmitted through physical platforms and electronic platforms. Completing a survey is a good example of the former and sharing data on social networks (e.g., Instagram photos, Yelp reviews, tweets) involves the latter. One example is the GoViral project (www.goviralstudy.com), a joint study between Harvard University and New York University that seeks to survey flu-like epidemics by requesting that sick people send mucus swabs for virus and bacteria analysis.

Second, there are data that humans generate passively. Take the case of mobile phones. When you load an app, do you know what data are being collected related to your activity? Most people are not aware about the kinds of data they are allowing the app to collect (Desouza and Simons 2014). Apps collect a great deal of data in the background while in use (and in some cases even when not in use, so long as the app is loaded on a device). There are also cases where individuals agree to share data passively; they acknowledge and are aware that data are being shared with an organization but do not actively engage in the sharing. The Center for Embedded Network Sensing at the University of California, Los Angeles, created the Personal Environmental Impact Report program that enabled voluntary participants to collect data on their carbon footprints (Mun et al. 2009). This was done through mobile data collection that included GPS location, which segmented data into trips; a lookup system for traffic, weather, and other context data; and an algorithm that calculated environmental impacts and exposure based on trips made.

Third, there are data collected by sensors on devices that the individual does not own or control. As you move through the streets in London, you are being monitored by cameras. Similarly, when you drive on most toll roads, you are being recorded. The City of Chicago has begun to pilot the Array of Things project (https://arrayofthings.github.io), which seeks to place 50 sensors throughout the city to collect data on various variables, such as the number of Bluetooth and Wi-Fi devices in the area and wind, heat, precipitation, and light intensity (Figure 4.2). Sensors will be located in boxes around the city with decorative shields to help them blend into the environment and will regularly capture data on resident actions.

Fourth, there are data that are generated by inferences made from allied data sources. For example, your iPhone might be able to tell us how much you walk per day. Similarly, your photos on Facebook and Instagram might provide insights into your eating behavior, and your credit card purchasing behavior might signal your health choices (e.g., you buy cigarettes). Through integration and analysis of these data, we can infer data on your health behavior and use this inference to customize insurance and healthcare offerings, for example. By combining data available from local governments (e.g., data on driver's licenses, tax and revenue, utilities), we can infer several outcomes such as financial viability or risk behavior of individuals.

Figure 4.2. Rendering of an Array of Things node (left) Array of Things air quality data collection system (right) (Array of Things)



SUBSPOTTING IN NEW YORK CITY

Citizens Daniel Goddemeyer and Dominikus Baur developed an app called Subspotting (www.subspotting-app.nyc) that captures, maps, and visualizes cell phone reception along the New York subway system. For the 660 miles of the subway, cell reception ranges from good to bad to ugly. On his regular commute, Goddemeyer noticed that reception was spotty, at best. He and his team used a prototype data logger to detect levels of reception for major cell phone carriers. He developed the app using this data that users could access at any time (Figure 4.3). The Metropolitan Transportation Authority even licensed and released posters with Subspotting information. As a result, New York residents can download the Subspotting app as well as view the posters to find out where they can get coverage in the subway.

Figure 4.3. Subspotting app (Subspotting)

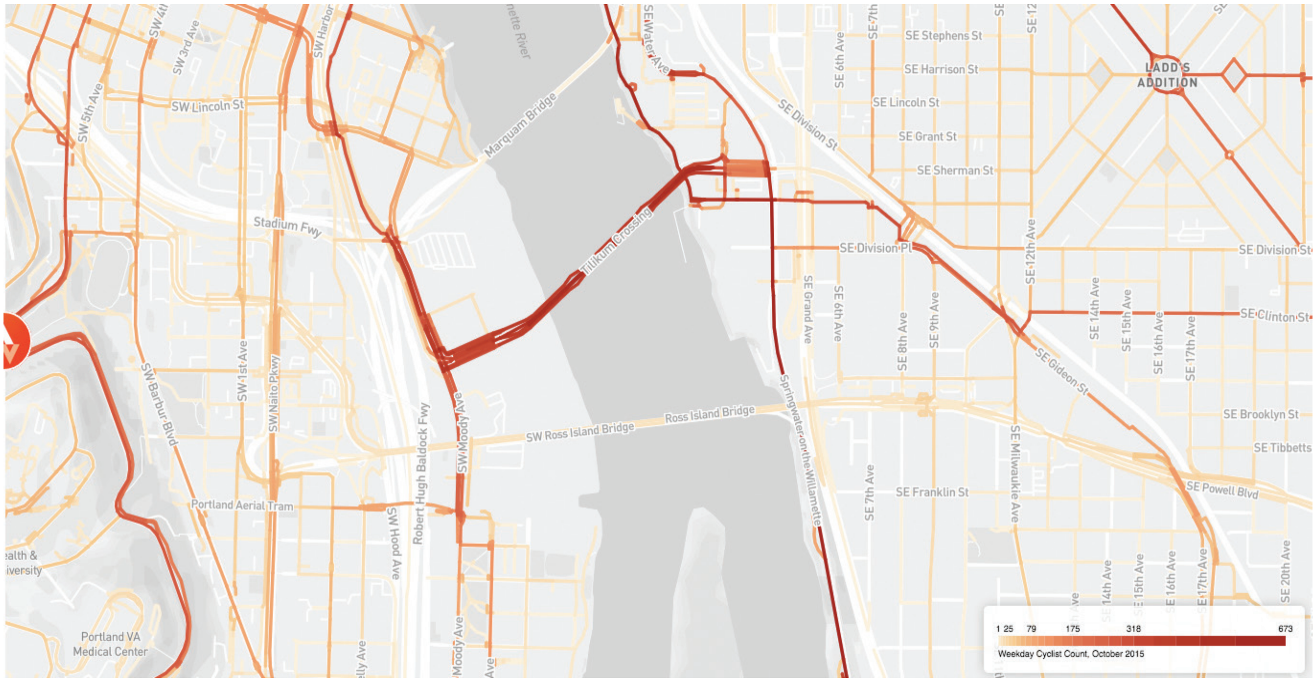


Figure 4.4. Visualization of Strava Metro data showing levels of bicycle traffic after the opening of the Tilikum Crossing in Portland, Oregon (Strava Metro)

differ in the extent to which they promote the sharing of data within organizational units or even the posting of data for the public to use (e.g., on open data sites and for events like hackathons). In addition, the licensing agreements outline the nature of the fees, the duration of the data provision, and the extent of data provided. As many private-sector organizations consider data to be their competitive asset, it is very important to ensure that you are working within your organization's data sharing and licensing agreements. Currently, there are not widely accepted principles for data sharing and licensing agreements in the public sector.

While there are a few forward-thinking cities and countries, more time and attention will have to be dedicated to documenting data-sharing experiences and expanding professional skillsets in data sharing (Smith and Desouza 2015). One example is the Scottish Data Management Board (www.gov.scot/Topics/Economy/digital/digitalservices/datamanagement), a group of officials that proactively design and manage data sharing, focus on reaping the benefits of data for society, and support the improvement of services by 2020.

The Oregon Department of Transportation (ODOT) is an example of an agency that purchased data from a private vendor, which was better data than it could have collected on its own. ODOT was the first state transportation agency in the United States to build a data feed that transformed planning, projects, and policy making through data-driven decision around bicycle planning (Maus 2014). ODOT sought to build bike lanes in multiple cities, but progress was constantly stunted by the unavailability of cycling data to locate the best and most used locations and routes. The agency looked to the popular cycling app Strava to collect geolocational data on cyclists. Strava captures over 400,000 bicycle trips per year showing where people are biking (e.g., streets, sidewalks) and how they bike (e.g., where they slow down or speed up) (Figure 4.4) (Maxwell 2015). ODOT purchased a one-year license from Strava for $20,000 to acquire datasets on cyclists. Using these data was much more accurate and efficient than previous methods of manually assessing these measures, which would have involved ODOT staff poring over traffic videotapes to count cyclists or having someone physically go out and watch for cyclists.

ODOT was able to identify patterns by mapping the datasets using its TransGIS tool to identify the locations that riders most used, assess the size and traffic density of those areas, detour planning during construction, schedule maintenance, and develop forecasting modeling of travel demand. The analysis of Strava's data, which covered five

TWITTER

Currently, there are 313 million active users monthly on Twitter (Twitter 2016). If you would like to tap into users' tweets, it will cost you. Data seekers have the choice of purchasing tweet data from Twitter Firehose, which provides access to all tweets, or one of Twitter's application programming interface (API) offerings. The Twitter Search API (https://dev.twitter.com/rest/public /search) allows access to data that already exist through search criteria such as keywords, usernames, and locations. There are limits to how much data can be collected: users usually can receive the last 3,200 tweets, and they are limited to the number of requests they can make. Unlike Twitter Search API, Twitter Streaming API (https://dev.twitter.com/streaming) is a collection of data put together in near real time. Users can set the criteria ahead of time, and tweets that match those criteria are gathered. Similar to Twitter Search, this service only provides a sample of tweets with the size of the sample varying based on the user's request and current traffic in the system. Twitter also sponsors competitions for access to its datasets. Individuals and organizations have to submit a proposal about how they might use data from Twitter, the problem they want to solve, and how social good will be advanced.

million biking miles traveled over a one-year period, resulted in the placement of three permanent bike counters and validated that particular areas needed planning or improvements. With this information, ODOT was able to make better decisions about which spots should be targeted for both auto and cyclist safety improvements. Additionally, ODOT, together with researchers at Portland State University, developed a smartphone app called ORcycle, which is a crowdsourcing app to record cyclists' trips and indicate locations where they experienced near collisions, crashes, or weather conditions that were unsafe—details that often go unreported (Figliozzi and Blanc 2015). Other cities are using similar crowdsourcing data collection strategies, such as the San Francisco County Transportation Authority's CycleTracks app (SFCTA 2016).

Evaluating Your Sources

Sources vary in the level of detail of data they provide and also the extent to which one might depend on them. For example, a human source, say a planning expert, might provide you very detailed information on a given issue. However, this person who has expertise may not be an affordable or regularly accessible resource. A social media or crowdsourcing platform will give you data that are much more representative when it comes to the views of digital natives but will probably not capture the sentiments of the adult population in your community. Sources also vary on the reliability of the data they provide over time.

Evaluating sources and the data they provide normally requires one to know about the methodology or process that the source employs to generate the data. Opening up the analytical process of a source may not always be possible. In these cases, sources are normally evaluated based on their track records of providing accurate and valuable data, which affects their brand recognition and authority. Sources also need to be evaluated on a temporal basis to ensure that they are still credible and valuable. Think of this akin to an expiration date on a carton of milk. On a regular basis, the planning unit must be alerted to evaluate their sources and update them if needed. Sources may move up or down in value or may even need to be dropped or reassigned to new domains depending on developments.

Ultimately, one must consider the costs and benefits associated with evaluating sources. The costs of evaluating sources are often amortized over time, if the source is used repeatedly for data. The cost of evaluating sources should not be prohibitively high, except in very rare cases (e.g., you do not have alternatives), and should always be less than the benefits one will get from the data.

In the past, organizations found it easy to attribute credibility to sources. These attributions were made based on either steps the data source took to get credentialed or certified or the source's official status (e.g., data from the US Census Bureau, the Internal Revenue Service, the National Weather Service). Today, we must find more innovative ways to ascertain the credibility of sources, especially those that are transient and have little history. One successful approach is to learn from crowdsourcing and open-source approaches to software development. The peers of individuals can provide ratings based on the contributions. Contributors self-select into various categories that represent their expertise, which also gives us a way to identify the domains of knowledge and competency.

In these platforms, credibility is earned and not assumed due to the position one held or the resources to which one has access. Individuals contribute elements to the project, and each element is evaluated for its value, usefulness, and quality. These ratings accumulate over time and are aggregated to ascertain the credibility of an individual within the community. A step further would be for individuals to share performance data on past projects and feedback on products with the community.

In addition to evaluating sources, one has to pay attention to the data put out by the source. While some sources retain their credibility over time due to the quality of data they provide, the data quality from other sources might not be consistent. In addition, a source that might provide high-quality data in one domain can provide low-quality data in another domain. In short, credibility is dynamic. It changes over time and even over short periods of time. Traditional sources, such as an expert from the early 2000s, might not be up to date with current data trends. Very little should be deemed credible by default, so planners must take extra steps to do their due diligence when it comes to evaluating data quality.

A critical issue for planners is the usability of data that are created anonymously or with little reference information. Data created from unknown sources can be problematic because the ability to evaluate the credibility of the data is paramount to the big data process. When this capability is diminished, use of these data becomes a critical issue. Even in instances where data are created from somewhat known sources, credibility can be difficult to gauge. For instance, attention has been placed on the value of crowdsourcing in finding solutions to problems. Some crowdsourcing efforts require identifying information about users, but others do not. If data are created anonymously, then their legitimacy and accuracy could be called into question because there is no one to hold accountable for the information. This is a challenge for planning organizations in light of a growing desire for and expectations of government transparency, the need to justify decisions, and the ability to ensure that the data used to make decisions are credible and reliable. Without being able to attribute information to a responsible party or understand the accuracy of information, planners are limited in their ability to use anonymous data. Planners and their organization have to decide how they will proceed with anonymously created data.

As a result of this difficulty, planners must seek ways to triangulate their data from multiple sources to boost credibility. Triangulation, or validation of findings through cross verification between two or more sources, is best when planned for prior to data collection. Data can be triangulated by looking at convergence from multiple sources. It is important to remember that often we are not looking for pinpoint precision (e.g., 93.2543) when working on most urban planning challenges. Rather, we are looking at ensuring that the data are accurate within given ranges (e.g., resident sentiment on an issue is above 90 percent positive).

Protecting Your Sources

With all of the interest in data collection and storage, the need to secure and protect sources has taken a backseat. It is important to remember that agencies have a responsibility to protect sources of data from undue harm and to ensure that the data collected are used for the intended purposes. In 2014 there were over a billion personal data records compromised; it was termed the “Year of the Hack.” These cases and more recent ones are driven by a few organizations with massive amounts of data: the billion data breaches from 2014 came from just 1,541 incidents (Kharpal 2015).

The protection of sources is extremely important because hackers are using new tactics to steal information. Before, hackers would try largely to steal credit card information but now they are looking at more long-term strategies. Hackers would rather steal a person’s entire identity so they can open whole new accounts that take longer to be detected. This means that risk exposure for individuals is expanding. Also, hackers are able to maneuver more covertly through systems than ever before. For instance, after a hacker cyberattack in 2014, the US Department of State was still trying to oust the hackers from its unclassified email network months later (GCN Staff 2015).

The continuing saga of the Internal Revenue Service (IRS) and the hacking of its systems affirm that the protection of sources is a serious and difficult thing to do. In 2015 the IRS was hacked and cybercriminals stole personal data from 333,000 taxpayers to file fraudulent returns. They ended up stealing $50 billion dollars from the US government, according to the Inspector General (Treasury Inspector General for Tax Administration 2015). Hackers in 2016 attacked the IRS again using malware to generate e-filing personal identification numbers so they could file more fraudulent tax returns using previously stolen social security numbers. Also in 2016, the National Security Agency (NSA) was hacked (Nakashima 2016). The hackers announced that they will auction “cyber weapons” made by the NSA—in effect, using the NSA’s information to work against it. While not on the same level, local governments also have extremely valuable information that can be hacked and abused by malicious agents.

To protect sources, three issues are critical. First, given the ease with which we can collect data at very granular levels and in real time, it is important to ensure that security measures are taken by individuals to protect their data and that they do not share sensitive data on platforms that are not protected. Individuals should not store sensitive data on these platforms because when compromised, they give hackers access to financial, email, and other pertinent accounts. The importance of this was illustrated when data storage site Dropbox was hacked in 2016 and a reported 68 million usernames and passwords were leaked (Gibbs 2016).

Second, when we collect data from humans, we are engaging in a compact, whether implicit or explicit, about how we intend to use the data. This is very important to remember, as individuals share personal information differently based on how they perceive it might be used. As a result, using data collected under an explicit compact to do X should not be used to do Y. Obviously there is no strict rule to determine X or Y, and certainly one might argue that Y is likely to be a slight derivation of the original X.

Third, when making data available on public forums to promote open data or for hackathons or other commercial uses, we must ensure that care is taken to anonymize the data. Anonymized data is extremely important for respecting individuals’ privacy, but it is not as straightforward as removing names or addresses. Consider open data and how, even when released anonymously, individuals can be subject to identification—such as when you combine open data from multiple sources. For instance, researchers at Carnegie Mellon University were able to detect social security numbers by exploiting open data with individuals’ place and date of birth (Acquisti and Gross 2009).

Fourth, you are only as strong as your weakest link. Given that in the big data world we are linking datasets

OPEN-SOURCE DATA AND COMMUNITIES

Complex technologies that planners would, at one time, have had to purchase are now available to them in open-source formats: open-source applications, technologies, e-services, and data. This increases opportunities to share more information through reusable, open-source technology and spend less time and money on technology development. Open-source communities are interesting models for us to learn from when it comes to evaluating sources. These are communities of independent individuals who work and participate in the development of technology. The abilities of open-source communities can be superior to those of any one organization because of the diversity of ideas and talents within the community, quicker market penetration, and faster innovation. Perhaps the most valuable aspect of open-source communities is that they shift the power of technology from information-technology companies to the information-technology user.

Although these communities are "open," they are not places where anyone can come in with no expertise. In many cases, users are rewarded for their efforts as they participate. For instance, Linux (www.linux.com) functions, in part, as an open-source operating system that allows any user to participate in coding challenges. Coders participating gain more opportunities and responsibilities as they successfully complete challenges. Rewards often include greater access to and control over a project. This ensures that the community is growing in the right way with the best and most trustworthy users. MIT's Urban Network Analysis, described in Chapter 3 (p. 50), is an example of open-source software available to urban planners that helps them understand and apply big data information related to cities and urban growth.

Individuals populate communities with technology-development skills. Key characteristics of these communities are transparency, trust, and clear outlining of projects. Transparency in the group allows everyone to know who has done what work, the process behind the work, and how much the work will cost. This helps to build trust, which helps participants in the community work together continually. Credibility among participants is also needed to build trust. Finally, clarity of the project process helps with quality assurance and moving the project from start to finish.

OBSERVATIONS.BE: WORKING FOR CREDIBILITY

Observations.be (www.observations.be) is a French-language online participatory (or crowdsourcing) monitoring platform of biodiversity in Belgium. The platform allows anyone who is registered to add entries for species they have observed (e.g., birds, mammals, insects, reptiles). Entries from users are what make the site work since user entries identify species on geospatial maps where photos, sound files, and commentary can be added. Another way Observations.be is structured is by collective action, meaning that working groups are central to the platform's function. Each group has an administrator in charge of validating each entry before it is integrated into the Observations.be database.

The Brussels Institute for Environment Management (IBGE) is the public institution overseeing energy and environmental issues. Due to insufficient people power, the IBGE is unable to maintain detailed databases of species monitoring. Over the last two decades, IBGE has been opening up monitoring efforts to social-sector organizations such as universities and nonprofit organizations; in fact, most IBGE databases are attributed to these organizations. The IBGE subsidizes Observations.be and integrates its database into its own database.

However, issues of credibility are prevalent since the government can only dispense information based on credible and verifiable data that are collected based on rules of science. As a result, Observations.be has protocols in place that enable validation of data. The validation of data focuses on correctly identifying a species and ensuring the credibility of aggregated data. In validating species data, the user is the first filter for incorrect information. A user who doubts the identity of a species can enter the observation in different categories that seem like the most accurate fits or add a question mark on the category line. In both cases, a species coordinator will connect with the user to help identify the species. The second filter is an automatic message that appears when a user enters a rare species asking the user to answer a series of questions confirming the sighting and to upload a photo backing up the claim. The third filter is a species coordinator (usually a veteran nature observer or specialist) who reviews the entry, contacts the user if there are questions, or sends an entry to a discussion forum of other veterans and specialists about whether to discard or accept the entry.

THE ETHICAL USE OF DATA

The pervasiveness of data and its use mean that new analytical and ethical questions need to be considered. Privacy concerns used to hinge on the concern of access to information, but now we need to be more concerned with how the information is being used. Data can be misrepresented, misused, and not safeguarded, to the detriment of the public.

For instance, following Hurricane Sandy in the Northeast, a study aggregated tweets and Foursquare posts to identify trends in New York in the days immediately preceding and following the hurricane (Crawford and Finn 2015). This study found that the day before Hurricane Sandy, grocery shopping peaked, while nightlife grew significantly the day after. Although these findings were interesting, they were skewed and could not be generalized to all of New York because less 20 percent of the US population uses Twitter and Foursquare and tweets received were largely from urban dwellers. The authors bring up several ethical issues related to using data that were collected from individuals who might have been at their most vulnerable and in need of help. Additionally, despite the information being posted online, the aggregation of multiple data feeds can generate intimate insights about individuals without their knowledge.

Another example is the Fitbit—a wearable monitor that captures information about users' activities as well as manual entry features for activities such as sex. The monitored activity populates user profiles that can be kept private or remain public (by default). The default public setting is meant to encourage social sharing and competitiveness among users. In 2011 it was discovered that some Fitbit users' sexual activities were actually showing up online in Google search results (Hill 2011). We suggest the development of an internal ethical framework to handle these issues.

across organizations, systems, and environments, we must take care to ensure that the entire data supply chain has necessary security measures in place. A chief concern for planners and municipalities should be hacking. Researchers found in Spain that smart meters were vulnerable to hacking (Ward 2014). A research lab found that roadside sensors and the data collected from them can be easily hacked on the streets of Moscow (Leyden 2016). These are examples of issues that must be handled before these technologies are deployed to prevent lapses in privacy and civil liberties—or even deaths.

Fifth, data on unauthorized platforms are being transmitted on unauthorized channels, such as sensitive information sent by personal email and transmitted through unencrypted servers.

PHASE 2: INFORMATION THROUGH ANALYTICS

We are now ready to begin to analyze the data in order to extract information. Much like traditional data analysis, we need to ensure that we conduct analytics with appropriate care so as to arrive at information that is accurate and valuable. The following discussion will review key practices to adhere to when analyzing big data specifically and applying analytical methods.

Data Cleaning

When data scientists and analysts receive data, the data will often be in formats that are incorrect, inconsistent, inaccurate, irrelevant, or incomplete. This so-called "dirty data" will require cleaning and correcting. Data cleaning is not optional, and this task is critical to the analytical process. Without it, insights derived from the data can be incorrect and misguide planning and public policy actions, such as financial or resource allocations or improvements to infrastructure.

In a *New York Times* article about big data, Steve Lohr explains that data scientists have to take on tasks including "data wrangling," "data munging," and "data janitor work"—the process of manually converting data from raw form to other forms that enable its use with analytical tools (Biewald 2015; Lohr 2014). Based on interviews and expert estimates, Lohr estimates that data scientists spend around 50 to 80 percent of their time collecting and cleaning data to prepare for analytics. That amount of time will likely remain high with the inclusion of other datasets from other sources to create unified datasets.

Data cleaning involves several activities, the most common being the following: (1) dealing with missing data (often times data records will have missing or incomplete information), (2) dealing with inconsistent data between systems (i.e., two different systems will differ in values of given fields), (3) dealing with inconsistent ways in which data on a single element are captured across systems (e.g., in one system eye color is noted by spelling out the word BLACK and in another by the abbreviation BLK), (4) dealing with errors in how data are recorded, and (5) modifying data formats and field structures to make them analyzable. Table 4.1 (p. 72) lists a range of additional data-cleaning issues that need to be addressed before the data are made available for analytics.

The following steps can help analysts improve and maintain data quality: (1) design a data entry interface that has database integrity constraints—such as data type checks, bounds on numeric values, and referential integrity—to prevent inconsistent data entries, (2) create organizational management structures that streamline data processes (e.g., systems that include data collection, archiving, and analysis) to limit error, and (3) automate data auditing and cleaning to correct errors (Hellerstein 2008). The emphasis in these steps is that value can be derived from any of these actions but that these actions should be prioritized based on organizational dynamics. Data-quality efforts can be organization wide (e.g., creating an organizational management structure that includes leadership and information technology staff) or simply a unit function (e.g., designing a data interface that allows for data integrity).

Data Integration

Once data from several sources are cleaned, they are then linked and connected so as to bring multidimensional perspectives and broaden situational awareness. The fusing of data is a nontrivial task that requires creativity and rigor. Creativity is important because identifying the fields that can be linked or connected in a database is not always easy. Hence, a fair degree of innovation takes place. This is especially true when trying to link databases that do not use the same fields consistently across the databases. Rigor is important to ensure that the data are not incorrectly manipulated or accidentally transformed during the integration.

Data integration and fusion require an understanding of how to link data across multiple systems and sources. For example, if you were to look at a database of drivers' licenses, you might be able to know who has access to a vehicle and a

TABLE 4.1. OTHER COMMON DATA PROBLEMS

Problem	Description
Illegal values (gender: Canadian)	Values outside of the domain range
Misspellings (Bosten, MA)	Typos or phonetic errors
Wrong references (gender: female instead of male)	Entry is entered incorrectly
Misfielded values (phone #: 73453)	Data entered into the incorrect field
Different value representations across sources (source1: marital status; source2: relationship status)	Value is represented differently across more than one source
Naming conflict (tired, sleepy, exhausted, fatigued, weary)	Homonyms where the same name is used for different objects or synonyms where a different name is used for the same object
Abbreviations or cryptic values (education: HS [high school or Hawaii State College])	Values that are unknown
Violated attribute dependencies (city: Providence, RI; zip code: 90210)	Values that should correspond but do not
Duplicates	Same entry listed twice (usually a data entry error)
Transpositions (name: Doe, J.; name: J. Doe)	Transposition of words (usually in a free-form field)
Embedded values (name: John Doe, 12345, 02/04/1978)	Multiple entries in one attribute field
Different interpretation of values across sources (source1: Euro; source2: dollar)	Value is interpreted differently across more than one source
Overlapping data (also object identity problem/duplicate elimination, merge/purge problem)	Problems with matching records referring to the same entity (person, object)
Uniqueness violation (John Doe, SSN: 12345; Jane Doe, SSN: 12345)	Values that should be unique to one person are attributed to more than one person

Source: Adapted from Rahm and Do 2000

permit to drive. This insight, while valuable on its own, can also then be linked to a database on taxes and revenues in order to understand the kinds of cars and homes of drivers and the taxes they pay, and if they pay their taxes on time. Furthermore, if these data are linked to the drivers' healthcare records, you might be able to learn about their overall health and specifics such as exercise or walking behaviors. Linking data across multiple systems and linking different quality-of-life indicators gives us greater situational awareness about objects and agents of interest to us.

Integration is a challenge because most analytic units create and operate datasets based on specific purposes or problems. This often results in disparate data silos that can further complicate data storage and data models being used, which creates challenges during the planning process. This can make cities data rich but information poor. In other words, many pieces of data are owned, but little is done with it to turn it into valuable information. Ideally, data captured through one source should be able to be easily integrated and linked to another. However, things are never that straightforward. Standards, rules, and regulations need to be developed about how to manage data use across organizations.

In 2007 the European Parliament established the Infrastructure for Spatial Information in the European Community (INSPIRE) (http://inspire.ec.europa.eu) directive to ensure that spatial data infrastructures are compatible and usable across member states. The main

CONNECTED VEHICLES AND SECURING DATA

Vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) capabilities can be used to alert drivers about accidents and traffic flows. Using V2V and V2I capabilities, cities can reduce traffic congestion, carbon emissions, fuel costs, and road accidents. *MIT Technology Review* listed V2V technology to be among the biggest technology breakthroughs in 2015 (Figure 4.5) (Knight 2015). The National Highway Traffic Safety Administration (NHTSA) has estimated that V2V technology could prevent over 500,000 accidents and save 1,000 lives annually if implemented in the United States. This technology is projected to be on the road as early as 2016. However, it has given rise to new types of privacy and security concerns. For example, in 2015 it was discovered that several Chrysler models could be remotely accessed through the internet. Hackers could access the critical functions of a car and conduct activities such as cutting the brakes and shutting down the engine.

Another issue is adoption and regulation. In 2014 the US Department of Transportation issued advanced notice of proposed rules for implementation of V2V technology in new model vehicles (NHTSA 2014). This means that the public sector and private industry will have to work together to make V2V technology work and be effective. Luckily, this has already begun. All major automakers have been sharing data on V2V technology to standardize communications between the different makes, models, and brands.

Another issue to contend with is privacy. The Federal Trade Commission has shared concerns about NHTSA's advance notice regarding GPS tracking and data collection as an issue of consumer privacy for drivers who do not want their travel data transmitted and who cannot opt out of the transmission of this data (Ingis, Signorelli, and Wolf 2015). NHTSA has noted these concerns and has committed to establishing comprehensive protections, which, undoubtedly, will have to be done in consultation with the automakers.

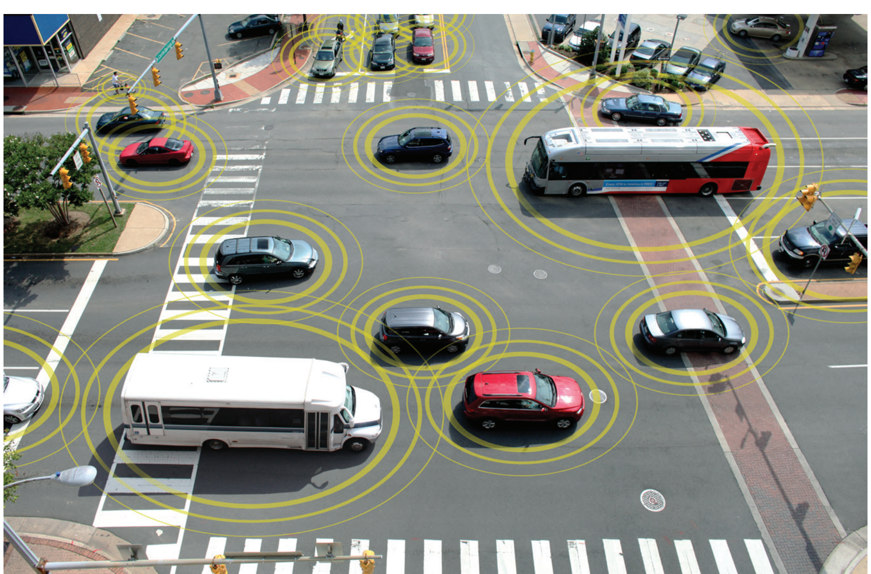


Figure 4.5. Vehicle-to-vehicle and vehicle-to-infrastructure technology (National Highway Traffic Safety Administration)

NEW YORK CITY PARKS AND RECREATION

The New York City Department of Parks and Recreation has a serious problem with trees. Trees have caused deaths over the years, and they are a hazard after serious storms. To combat this, the department began a preventative maintenance program that involved regularly scheduled pruning and grooming of large trees to reduce damage and risk. For a number of years, it kept various records on city trees, such as a 2005 tree census, planting records from 2005 to 2011, block-by-block pruning data in Brooklyn, work orders and requests, inspection records, and an earlier tree management database that tracked 311 forestry requests, work orders, and requests from 1998 to 2008 (Data Kind 2012). Using these years of data, the department wanted to know causally if this maintenance reduced hazards and damages the following year.

Despite having a great deal of data, analytics could not be easily run to answer the department's question because of the way the data were collected. The data had been gathered for reporting purposes and not for linking to other datasets, a feature which would be necessary to answer a causal question. There was inconsistency with the unit of analysis—trees had no unique identifier, and tree-pruning data were collected at the block-by-block level while other tree-cleanup data were collected on an address level. To answer the question of the program's effectiveness, analysts had to settle on a unit of analysis; they chose to collect data at the city-block level. Different analysis tools were used, including Python to pool data from disparate sources and Quantum GIS (open-source GIS) and CartoDB (mapping system) to create data visualizations. An analyst merged, cleaned, and analyzed the data—a process that took hours—and ran models that found tree pruning reduced by 22 percent the number of times the department had to send emergency cleanup crews (Olvasrud 2013).

reason for this concerted effort is to make all data usable so complex issues can be understood and addressed. INSPIRE requires member states to establish and adopt common implementing rules in the following areas: data specifications, metadata, data and service sharing, network services, process and procedures, and monitoring and reporting. A commission of representatives from the member states oversees the setup for data integration and establishes binding regulations and rules. It does not require the collection of new data but rather that data only be collected once and stored in the appropriate format (Iannucci, Congedo, and Munafò 2012).

Data Analysis

A wide assortment of approaches is available to analyze data; several of the common approaches were highlighted earlier in this report. A critical ingredient for analysis is to have human capital in place for conducting these analyses and building routines and models that can be deployed based on certain rules. McKinsey & Company projected that by the year 2018 there will be a shortage of 140,000 to 190,000 data scientists (Manyika et al. 2011). Hiring data scientists is not inexpensive (according to Glassdoor, data scientist salaries in the United States range from $76,000 to $148,000 with a national average of $118,709) and, given their scarcity in the marketplace, one can expect them to be a costly resource until workforce needs are addressed by educational institutions and training programs.

However, the opportunities for success are extensive if the right data scientists are in place with the right support. For instance, transportation networks are extremely important for travel, commerce, and personal pleasure but issues like congestion and infrastructure often impede this at big costs. At the Brasilia International Airport in Brazil, airspace congestion is an expected issue as the population continues to grow, and this congestion is currently a concern. Officials have introduced a new system that uses GPS data to optimize air space. The system assigns each plane its own flight path; these calculations require vast amounts of data about the distance, speed, and capabilities of an aircraft to find the shortest flight path. As a result, this decreases delays in the current system of aircraft queuing, in which planes wait behind one another to land. After the first deployment of the new system, the airport was saving 7.5 minutes and 77 gallons of fuel per landing (Neumann 2015).

Although complex at times, data analytics does not have to be left up to the data scientists only. Communities of practice are communities within a profession that develop knowledge contributing to the professional body. This can be individuals from across an organization who want to contribute to data-based activities. In Australia, a national data analytics center will operate its own community of practice that is open to all personnel who are interested in data analytics (Gardiner 2015). Nontechnical communities working with data benefit analytics because different perspectives can strengthen insights about which data to include. The US government similarly hosts a community of practice comprising various government officials at different levels to use data analytics and data sharing to overcome challenges (www.gao.gov/aac/gds_community_of_practice/overview).

Several other choices are available for bringing in expertise to conduct analytics. A popular method of analytic tool building is through community collaborations with civic entrepreneurs through data competitions, hackathons, and crowdsourcing. Data competitions are sponsored competitions that pose a problem to a body of groups and individuals who compete to provide solutions. In the case of big data, these events are technology themed and participants are usually using open data. The benefits of hackathons are that they focus on solving local challenges, collaborations are fostered across many disciplines and organizations, and platforms are created for producing creative solutions to civic glitches. Hackathons are event based, last a few days, and are issue focused.

Crowdsourcing platforms, on the other hand, have more persistence. These platforms allow individuals to compete and collaborate on various competitions and challenges. There are two forms of crowdsourcing: (1) where individuals come together to share information and opinions (known as crowd wisdom, crowd review, or crowd funding) and (2) where people come together to work on data challenges and data analysis for the social good (known as crowd creation or citizen science) (Table 4.2, p. 76). An example of the latter is Foldit (https://fold.it/portal), an online puzzle-solving site that allows users to contribute to scientific knowledge by engaging in a game that lets them work with the structure of proteins. Finding the best way to fold proteins can help prevent or cure illnesses, and users' competitive play helps with protein design.

Galaxy Zoo (www.galaxyzoo.org) is a web-based citizen science project that uses citizen volunteers to help researchers manage the increased flood of data on the formation of galaxies. It is one of the most well-known citizen science projects in the world. The key way to learn about the formation and history of a galaxy is from the shape of a star in the galaxy. Galaxy Zoo uses images acquired from the Sloan

TABLE 4.2. CROWDSOURCING PLATFORMS

Category	Description	Example
Crowd wisdom	Technology-enabled crowdsourcing initiatives to solve innovation challenges or complex problems	• Innocentive (www.innocentive.com) • Chaordix (www.chaordix.com) • NineSigma (www.ninesigma.com)
Crowd creation	Technology-enabled crowdsourcing initiatives to produce marketable intellectual goods	• Threadless (www.threadless.com) • 99designs (https://99designs.com) • Naming Force (www.namingforce.com)
Crowd review	Technology-enabled crowdsourcing initiatives to promote knowledge sharing and combination for decision making	• Peer-to-Patent (www.peertopatent.org) • IdeaScale (https://ideascale.com) • Consensus Point (www.consensuspoint.com/)
Crowdfunding	Technology-enabled crowdsourcing initiatives to support the raising of funds for projects, businesses, or initiatives through the collective effort of many	• Kiva (www.kiva.org) • GrowVC (www.growvc.com) • Indiegogo (www.indiegogo.com)
Crowd democracy	Technology-enabled crowdsourcing initiatives to promote open government	• Patient Opinion (www.patientopinion.org.uk) • Open Ministry (www.openministry.info) • Data.gov.sg (www.data.gov.sg)
Citizen science	Technology-enabled crowdsourcing initiatives to promote involvement of nonscientists in research projects	• Citizen Science Alliance (www.citizensciencealliance.org) • PatientsLikeMe (www.patientslikeme.com) • Cornell Lab of Ornithology (www.birds.cornell.edu/citsci)
Citizen journalism	Technology-enabled crowdsourcing initiatives for sharing and aggregation of content of journalistic value	• CNN iReport (www.ireport.cnn.com) • Demotix (www.demotix.com) • AllVoices (www.allvoices.com)
Crowdsourcing for crisis response	Technology-enabled crowdsourcing initiatives for involvement of individuals in disaster/crisis response and recovery	• Social Media for Emergency Management (www.sm4em.org) • Humanitarian Tracker (www.humanitariantracker.org) • HealthMap (www.healthmap.org)

Source: Smith, Ramos, and Desouza 2015

Digital Sky Survey where volunteers can go in and classify images. Galaxy Zoo was launched in 2007, and within one hour on its first day it received almost 70,000 classifications. During its first year, it received 50 million classifications by more than 150,000 volunteers (Galaxy Zoo n.d.).

Another example is Challenge.gov (www.challenge.gov/list), a crowdsourcing platform started by the US government that gives citizens opportunities to develop solutions for particular public problems. Individuals or groups use data and open data to develop websites or applications. Challenge.gov partners with the private sector for judging and offers prizes for the winning solutions. Challenges include creating technical proposals for items such as wearable alcohol biosensors and plans to help consumers block robocalls. Prizes can include money and other special incentives, such as mentorship and engagement with the sponsoring agency. Ten data scientists from the nonprofit organization Data for Social Good worked with the Chicago Department of Public Health to build a model that predicts the risk to a child of lead poisoning. The most common approach to detecting lead poisoning is through blood tests of people who have already been exposed. To intervene before this happens, the team used two decades of data from blood tests, home inspections, property value assessments, and the census to help inspectors prioritize homes that have been listed as potentially hazardous and identify children with high risk of lead poisoning (Potash et al. 2015).

Building Analytical Tools

A critical aspect of the analysis stage is having tools to support the analysis. Analytical tools from two sources are available. First, a wide assortment of technology players has tools to offer for all types of analytical approaches (traditional

MOOCS AND ANALYTICS

For planners looking to develop skills in analytics, MOOCs, or Massive Open Online Courses, can offer skills and trainings that are sometimes free. MIT offers a course called The Analytics Edge (www.edx.org/course/analytics-edge-mitx-15-071x-2), a free, 12-week online course for people interested in improving or learning analytics skills. Stanford University offers Machine Learning (www.coursera.org/learn/machine-learning), an 11-week course about machine learning analytic techniques; the course has a fee, but financial aid is available. We encourage planners to use these resources to acquire baseline knowledge about the intricacies of the various analytics methods and new trends in the field of data science. Table 4.3 provides a detailed list of many MOOCs of interest related to analytics.

TABLE 4.3. BIG DATA AND PLANNING MOOCS (MASSIVE OPEN ONLINE COURSES)

Data & Analytics	Smart Cities	Urban Planning
Class: Framework for Data Collection and Analysis **Hosted by:** University of Maryland, College Park **Link:** www.coursera.org/learn/data-collection-framework	**Class:** Smart Cities **Hosted by:** ETHx **Link:** www.edx.org/course/smart-cities-ethx-ethx-fc-03x	**Class:** Sustainable Urban Development: Discover Advanced Metropolitan Solutions **Hosted by:** DelftX WageningenX **Link:** https://www.edx.org/course/sustainable-urban-development-discover-delftx-wageningenx-ams-urb-1x
Class: Graph Analytics for Big Data **Hosted by:** University of California, San Diego **Link:** www.coursera.org/learn/big-data-graph-analytics	**Class:** Introduction to Metrics for Smart Cities **Hosted by:** IEEEx **Link:** www.edx.org/course/introduction-metrics-smart-cities-ieeex-scmtx-1x	**Class:** Urban Water—Innovations for Environmental Sustainability **Hosted by:** University of British Colombia **Link:** www.edx.org/course/urban-water-innovations-environmental-ubcx-water201x
Class: Combining and Analyzing Complex Data **Hosted by:** University of Maryland, College Park **Link:** www.coursera.org/learn/data-collection-analytics-project	**Class:** Health in Smart Cities **Hosted by:** IEEEx **Link:** www.edx.org/course/health-smart-cities-ieeex-schc-x	**Class:** Quality of Life: Livability in Future Cities **Hosted by:** ETH Zurich **Link:** www.edx.org/course/quality-life-livability-future-cities-ethx-fc-02x-1
Class: Data Visualization **Hosted by:** University of Illinois at Urbana-Champaign **Link:** www.coursera.org/learn/datavisualization	**Class:** Big Data for Smart Cities **Hosted by:** IEEEx **Link:** www.edx.org/course/big-data-smart-cities-ieeex-introdatax	**Class:** Re-Enchanting the City: Designing the Human Habitat **Hosted by:** University of New South Wales **Link:** www.futurelearn.com/courses/re-enchanting-the-city?utm_campaign=Courses+feed&utm_medium=courses-feed&utm_source=courses-feed
Class: Big Data Applications and Analytics **Hosted by:** Indiana University **Link:** https://bigdatacourse.appspot.com/preview	**Class:** Smart Cities—Management of Smart Urban Infrastructures **Hosted by:** École Polytechnique Fédérale de Lausanne **Link:** www.coursera.org/learn/smart-cities	**Class:** Designing Cities **Hosted by:** University of Pennsylvania **Link:** www.coursera.org/learn/designing-cities

Source: Kevin Desouza and Kendra Smith analytics tools include SAS, SPSS, Minitab, and Matlab). Off-the-shelf solutions are available and can be installed with ease, but they are not cheap; the cost of licenses can be upwards of $10,000 per year, per user. For example, a single-user license for IBM's SPSS Modeler costs $4,530 for a personal-use license, $6,800 for a professional-use license, $11,300 for a premium perpetual license, and $5,000 for a social media analytics (cloud-based) license (IBM 2016). Most of the time the off-the-shelf systems will need to be customized to be able to analyze data and connect to data streams of interest. This is where further costs need to be accounted for in the form of consulting services. Most commercial vendors will gladly reduce the price of their technical solutions, and some might give them away for free, but you often end up paying for consulting services.

Second, thriving open-source communities are building analytical tools. These tools are made available to the public in the form of source code, technical manuals, and user communities that work on next iterations of these solutions, share feedback, and fix and report bugs. The advantage of these solutions is that they are normally free or available at a nominal charge. The challenge with these tools is that they are being developed by technical communities, and users often need to have a high degree of software engineering knowledge to be able to install and use them. These solutions may also be as unstable as those provided by traditional vendors. Unfortunately, users do not have much recourse (contrary to a traditional breach of contract) as these are provided as is.

Sharing and Reusing Analytical Routines

Given the time and expense involved in building and customizing analytical tools, it would be a shame if these resources were not shared and reused. Analytic reuse can provide greater insights and actionable results because information is timelier; it benefits more than one aspect of the process to gather insights. Effective reuse requires that data requirements be isolated and integrated into the analytic process. This enables effective and efficient access to information. Also, with the reuse of analytical routines, next steps become more predictable.

To do this, good governance is needed to store the various analytical routines and capture the most current and successive versions of the tools. This requires sufficient context on the analytical routines in the form of detailed documentation that accompanies systems and tools, but also shorter indexable descriptions that will enable search and retrieval of the tools. The costs need to be amortized over successive uses to justify the upfront investments made.

Further, experts or builders can connect with and ask questions about given tools before they need to use them. This guarantees that the tools are being used appropriately and care is being taken to ensure that the data being analyzed are suitable for the tools. A mechanism should also be in place for users to share feedback and comments about their experiences with the tools. Finally, there should be an overall repository known to the organizational members where the tools can be stored, and on a routine basis the repository should be curated to ensure that older versions of tools are archived, new tools are featured, and user feedback is addressed.

Corridor Simulation Modeling (CORSIM) (www.dot.state.mn.us/trafficeng/modeling) is a comprehensive simulation program that is used in Minnesota to model traffic and traffic control systems using commonly accepted models for vehicle and driver behavior (McCarthy 2005). A network of practitioners around the state were using CORSIM and developing ideas and models but those ideas and models could not be readily reviewed, understood, and shared by others due to a lack of standardization. In response, the Minnesota Department of Transportation developed an advanced CORSIM course and training manual to aid with application of the CORSIM software. Specifically, the manual standardized methods and computer coding in the network. This enabled the sharing and reuse of tools with significantly decreased review times and improved performance of users.

PHASE 3: INTERPRETATION OF ANALYTICAL OUTPUTS

We are now ready to make sense of our analytical results. This phase is all about ensuring that we are interpreting our analytical results in the most accurate manner and within the appropriate context. Analytical outputs can easily be misinterpreted both accidentally and purposefully to advance an agenda. It is important that care be taken so as to not fall into common traps when interpreting analytical outputs.

First, interpretation of results in a timely manner is essential. Insights that are outdated and no longer useful are a waste of time for all parties involved. Setting up organizational processes that make data interpretation a regular part of operations is essential to making timely interpretations. Second, the right people should be engaged in the interpretation process. These can be individuals internal to the organization, outside experts, and the public

VOLKSWAGEN AND ANALYTICAL OUTPUTS

Automaker Volkswagen was caught manipulating software in some of its diesel vehicles in order to cheat emissions tests in Europe and the United States. US regulators found that Volkswagen installed "defeat" devices (or parallel instructions) that modified emissions control systems when testing was sensed and signaled lower pollution levels. When not in testing, the vehicles were exceeding legal emissions levels by 40 times higher than allowed. The lower emissions levels made Volkswagen eligible to receive "green car" subsidies and tax exemptions in the United States. The lower emissions ratings also entitled drivers to federal tax credits for emissions reductions.

A research team at West Virginia State University conducted a study of two Volkswagen vehicles and one BMW and put them on the road (as opposed to a dynamometer, or "car treadmill") along five different routes with varying traffic levels and terrain. They found that the cars had the same emission levels along all parts of the routes—whether the Volkswagens were idling in traffic or going 75 miles per hour, the same high emission levels were present (Glinton 2015). The research team gave their results to the US Environmental Protection Agency (EPA), which began investigating the issue. Volkswagen heavily denied any problems with the devices and the vehicles and blamed the researchers. The EPA studied the issue and found the defeative device. Using analytics, it set the travel parameters in the vehicles to indicate a regular road trip rather than a testing scenario. As soon as the testing began, the vehicles began releasing high volumes of nitrous oxide.

(usually through crowdsourcing activities). Employees inside the organization have internal knowledge and expertise that is helpful. These are people who are likely to have grown within the organization, have a vast amount of organizational knowledge and understanding, and will likely influence the future of the organization. The value of internal knowledge is immense; it provides important organizational context. This context brings to bear specific organizational considerations that are important for contextualizing findings. This knowledge also ensures that we take generalized findings and apply them in a manner that is culturally sensitive. However, when relying on internal knowledge, it is important that the following do not occur: the emergence of groupthink, the manipulation of interpretation to support predisposed positions, or the development of incentives to willfully misinterpret or ignore interpretations.

External consultants are individuals who have significant expertise in an area but who do not have organizational ties (and so this does not include individuals internal to an organization with expert knowledge). In theory, consultants bring an unbiased opinion to interpretation, and they can help see opportunities and threats that individuals with the organization cannot recognize. They also have broad expertise as they have seen how things play out across a range of organizations and situations and can provide a broader perspective. However, the use of consultants should be tempered. James Shanteau, a scholar of expert decision making, has found that experts can be inaccurate, unreliable, biased, and lacking self-insights. He cites the psychological desire to see experts with an "aura" or mystique not possessed by others, which instantly makes their opinions more valuable (Shanteau 1988).

The public is the least-used source for data interpretation. When the public is used, it is traditionally through citizen engagement platforms that are attempting to leverage the "wisdom of crowds," or the collective intelligence of groups to create new solutions and opportunities. Citizen intelligence platforms are networks that enable communities of everyday citizens to work toward solving complex social problems and seizing opportunities for innovation. These platforms come in all shapes and sizes and many of their data are making their way into sophisticated data analyses. CoUrbanize (www.courbanize.com) is a collective intelligent city development platform started by a city planner to provide a way for citizens to engage in the development of their neighborhoods. The platform distributes information and allows users to give feedback through questions, advocacy, and comments.

However, the wisdom-of-crowds approach can yield poor results that in turn can lead to the foolishness of crowds and subsequently bad data. A study by a Swiss scientist suggests that modern technology actually makes it more difficult to benefit from collective intelligence. Researchers placed 144 college students in isolated cubicles and had them answer questions with a factual answer (Lorenz et al. 2011). The individuals were generally correct as a crowd. They then gave their subjects access to the other answers and, as a result, the students adjusted their estimates based on the crowd feedback. The range of answers changed and the researchers found that the students were imitating each other, which ended up magnifying biases and led to worse guesses.

To curb the negative effects of interpretation, multistep processes to interpret insights are helpful, and they can include a blend of expert, public, and internal information. One might begin by using the internal staff to make an initial assessment of what the data are telling them. They might then solicit experts and even the public to comment on the findings and provide suggestions for revisions and alternative explanations.

Sharing Analytical Outputs

The manner in which analytical outputs are shared is an important decision. Outputs can be shared via reports that describe the results, provide context, and examine the pros and cons of the recommendations or outcomes. Visualizations are a method of sharing analytical outputs using platforms such as Tableau (www.tableau.com) and Esri (www.esri.com). They help create awareness and provide valuable insights into how a city is performing. There are many different visualization types that run the gamut from static visualizations, such as infographics, to dynamic and interactive ones, such as advanced heat maps.

Visualization through map illustrations has been a very important way to convey data in urban planning. Visualizations traditionally come after simulation models are run; visualizations reveal the results of the models. For instance, the Municipal Art Society of New York developed a simulation tool that shows where development could bring the most change in New York. The tool is called Accidental Skyline (www.mas.org/urbanplanning/accidental-skyline) and it allows the user to go to a specific borough and parcel of land to see where there are available air rights (rights that developers quickly snatch up in order to build tall buildings) (Municipal Art Society of New York 2016). The tool essentially shows both where future development could happen and where other development has already occurred.

Decentralized energy has emerged as a technique to lower carbon emissions and transmission losses by producing energy closer to where it will be used rather than at a large plant. While high-density areas are plagued with increased carbon emissions, smart-city technology can provide options to lower carbon energy by combining power and heating systems and connecting them to district heating networks. With the goal of delivering 25 percent of London's energy supply through decentralized energy by 2025, the mayor of London created a program to identify opportunities through heat mapping and energy master planning (Greater London Authority 2016).

The London Heat Map (www.london.gov.uk/what-we-do/environment/energy/london-heat-map) provides spatial information on fuel consumption, carbon dioxide emissions, energy consumers, energy plants, and heat density to show opportunities for decentralized energy. Map information also includes geographic information (e.g., coordinates, postal codes, addresses), ownership, topology, fuel sources and heating supplies, primary energy consumption, number of dwellings, gross internal floor area, installed power, thermal capacity, year of construction, and new developments' start and completion dates. These data are regularly updated and the map is fully interactive.

Visualization tools can send a variety of messages; the main message is found not just in what the visual conveys but also in the degree of difficulty of the interaction. People can be given access to the data behind the visualizations through interactive models where they can play around with data, alter scenarios, and make decisions based on their experiences with the model. For instance, the *New York Times* Budget Puzzle (www.nytimes.com/interactive/2010/11/13/weekinreview/deficits-graphic.html) is an interactive visualization of public spending. In 2012 the Pentagon committed to a $450 billion strategy to reduce spending over 10 years and possibly even more if Congress enacted more reductions. The budget tool allowed users to make their own budget strategy for how they would reduce spending. Users could see how adding each line item changed the budget and by what percentage. The exercise exemplified just how difficult it is to make these cuts.

Importantly, no matter what the format of shared analytical outputs, context should always be included. Results can only be interpreted if they are in the proper context. Any caveats with the data or results need to be disclosed so there is not a wide misinterpretation of data. Further, the full data set can be made available, if possible, because in some cases people want to replicate the results to ensure that they are valid and accurate. This might necessitate someone in the agency being the key contact for anyone with questions about the data. While having a key contact person might seem simple to implement, in many cases requests for more information about data go unheard (our own experience with requests for more data for our own research projects confirm this). Further, replication, context, and verification of data are important in instances of data manipulation. All data are not good data, or are even what they seem to be—so users must do their due diligence before taking on big data and analytics.

Feedback on Interpretations

After analytics have been shared, researchers should solicit feedback on the interpretations. This includes critical questions, suggestions, and ideas about future work. It is important to understand the implications of the knowledge in question before constructing a course of action. Feedback can be gathered from citizens, individuals internal to the organization, and outside experts. Citizen feedback can happen in town hall or community forums or through websites with comments sections or citizen engagement platforms. If community feedback is being gathered, a dedicated person should be monitoring the conversation and engaging as an active participant in order to keep conversations on track and prevent them from being hijacked by individuals who have other motives and goals.

Acquiring feedback about the analysis is important to outcomes because there are many places to land on a decision-making spectrum with the two extremes of intuition and overanalysis. Decisions made based on hunches, feelings, and guesses are likely to be mistakes (Langley et al. 1995). Conversely, decisions made by overanalyzing a problem and the data are likely to prevent timeliness. The key is to find a balance between the two. Citizinvestor (www.citizinvestor.com) is a crowdfunding and civic engagement platform that allows government entities to post government projects on the site for funding. The site allows citizens to invest in their communities and become empowered. Feedback is offered through citizens' willingness to invest in the projects. While this is a rather simplistic way at looking at decision making, it certainly provides feedback on an issue. Not all communities can fund their own projects, but even page visits and written feedback help move decisions forward.

Further, platforms accomplish another important task for communities: they spur innovation and do it more quickly. Erik von Hippel developed the concept of lead users, or individuals that experience a specific need before the general marketplace, which positions them to benefit in their attainment of a

IMPLEMENTING ANALYTICS-BASED POLICY IN ISRAEL

Analytics is helping solve congestion problems and lowering pollution in Tel Aviv, Israel. Tel Aviv and Israel have some of the highest levels of traffic density per mile relative to other European and Middle Eastern cities and countries. In 2009 Israel instituted a "green tax" program that subsidized the purchase of cleaner cars by making individuals pay based on the amount of vehicle pollutant emissions produced by their vehicles. In 2011 the city began using a smart traffic management system to capture the density and congestion of vehicles in real time. On a 13-mile fast-lane highway, a dynamic toll system calculates fees based on current traffic and demand. A video-based license plate recognition system detects drivers and counts the number of cars on the road. If traffic density is high, then toll prices rise; if there are fewer cars on the road, then fees are lower. Toll charges are adjusted to the minute, and the amount charged is displayed on message billboards. The dynamic toll system helps reduce congestion and, as a result, carbon dioxide emissions (OECD 2016).

However, while Israel's green tax has been heralded as a success, it has had some unintended consequences. A 2016 report by the Organization for Economic Co-operation and Development showed the program had a "tremendous" impact; by 2014, approximately 83 percent of cars being sold had the lowest pollution grades compared to 19 percent in 2009 (OECD 2016). However, the program also had the unintended consequence of increasing car sales thanks to the government's subsidizing of car purchases, which effectively made cars much cheaper. The increase in car sales increased traffic congestion, which also raised pollutant emissions (OECD 2016).

THE TROUBLE WITH IMPLEMENTING BIG DATA ANALYTICS

While many big data initiatives are being implemented around the technical issues related to big data, little attention has been paid to the value created through big data and analytics. The benefits of big data have not been fully realized in the public sector, and conditions must allow data-driven innovation to flourish. Daniel Castro, director of the Center for Data Innovation (www.datainnovation.org), asserts that "most of the policy debate in Washington has been on how to minimize potential harms from data, especially around privacy, rather than on how to enable more and better uses of data" (Castro 2014).

Further, there are ethical and moral dilemmas around implementing policy based on predictive analytics, especially in matters of quality of life, safety, and liberty. For instance, predictive policing is a popular trend in some cities where crime has decreased as a result of efforts that either stop or anticipate criminal behavior before it happens.

While successful, translating these efforts into policy presents serious ethical dilemmas, such as perpetuating the profiling of and discrimination against individuals in lower socioeconomic groups or becoming a stopgap measure that decreases opportunistic crime but not the root causes of crime. For example, the Chicago Police Department uses a method of predictive policing that focuses on people (as opposed to geographic areas) who are most likely to commit a crime. Called the "Heat List," the system contains names of 400 people most likely to have some involvement in a violent crime. This person-centric form of predictive policing has the potential to result in race, class, and gender profiling solely on the basis of past behavior (Moraff 2014).

LONDON CONGESTION CHARGE: THIRTEEN YEARS LATER

In 2003 London became the first major city in the world to introduce a congestion charge program to reduce the flow of traffic into and around the city center (Stockholm, Singapore, and Milan have since followed). When cars enter the Congestion Charge Zone (CCZ), which is indicated by markers on streets, drivers are charged a standard fee of £11.50. The aim of the CCZ was to reduce traffic levels in the central area and invest the funds collected into London's transportation system. During the first 10 years, the CCZ raised £2.6 billion with about £1.2 billion (46 percent) of net revenue invested in public transportation, road and bridge improvements, and walking and cycling infrastructure.

Data from the program has helped with evaluation of the policy and future directions. For instance, after the CCZ program reached its 10-year mark, analysis showed it reduced the number of vehicles entering the charging zone by 23 percent, making room for improvements to the urban environment through increased road safety measures and prioritized public transportation upgrades. Also, the reduction in traffic flow was making some difference, but problems with nitrous oxide levels continued.

This fact gave birth to a new policy: the Ultra Low Emission Zone (ULEZ) (Transport for London 2016). The ULEZ plan has focused on the actual toxic fumes coming from cars rather than solely trying to reduce traffic in an area. It would require all private motor vehicles to meet standards higher than the Euro standard for emissions, or drivers will face daily charges ranging from £12.50 to £100. The ULEZ will not be fully functional until 2020, giving drivers time to adapt to the new compliance standards. It was the evaluation through data of the CCZ policy, its implementation, and its impacts that helped London officials decide on the next steps as they proactively seek to lower emissions.

solution that meets their needs (Von Hippel 1986). The car company Local Motors (https://localmotors.com) used a lead-user method of development to co-create vehicles through an online platform and a virtual community of car enthusiasts, engineers, and designers. They developed the Rally Fighter, the first vehicle in the world to be produced by this method of co-creation, which was featured in the film *Transformers 4: Age of Extinction*. Research has shown that this methodology has significant chances of commercial success because of lead users' propensity to identify needs that will become commonplace for most users as their usage of the product increases. Lead users also modify, customize, and enhance the existing product through the creation of add-ins and workarounds (Morrison, Roberts, and Midgley 2004; Schreier, Oberhauser, and Prügl 2007; Urban and Von Hippel 1988).

PHASE 4: DESIGN AND IMPLEMENTATION OF EVIDENCE-DRIVEN POLICY

Once we have made sense of the analytics, we are ready to begin the process of considering various policy options and designing strategies to implement outcomes.

Experimenting with Policy Outcomes

When possible, it is advisable to run experiments to test the efficacy of policy options and to study both the intended and unintended consequences of courses of action. Experimentation is a process that planners may not be comfortable with, but it is important to understand how you might go about implementing solutions. Implementation of a solution can take many forms, and often there are tradeoffs involved with each possible trajectory. Experimenting might allow you to see how well you witness the intended consequences and to identify some of the unintended consequences. Experimentation is risky, as it might reveal things that warrant going back to the drawing board and beginning again. However, experimentation always leads to learning, which is vital. There is no such thing as a "failed experiment." Even when you do not get the results you had hoped for and there is not support for a hypothesis, this is an opportunity to learn. Experimentation also allows you to test out different options and to study how theoretical ideas play out when they are implemented. Experimenting allows you to take full advantage of data and ensure that you are truly working toward evidence-driven policies.

In the private sector, experimentation is common in the implementation stage. For example, Subway is a company that has used analytics to drive success in its 40,000 stores both domestically and internationally. The company uses analytics through a loyalty program and electronic subscriptions to promotions. It found that when analytics programs were geared toward maximizing sales, it would promote to customers based on what they ordered most often. However, this was found to be destructive in the long term because the company found that expanding a customer's sandwich repertoire was a better strategy—the more sandwiches customers like, the more likely they are to be loyal return Subway customers (Horst and Duboff 2015).

Today, the use of randomized control trials (RCTs) is growing in popularity. RCTs are scientific experiments in which individuals or groups are randomly given one or another treatment. This is rigorous experimentation methodology because it minimizes bias and balances known and unknown factors. RCTs are viewed as one of the best methods for producing a standard of evidence of the effectiveness of an intervention. One example is the randomized trials researchers at the Abdul Latif Jameel Poverty Action Lab (J-PAL) use to answer critical questions about poverty and related policies, in which they create comparison groups to measure outcomes and effectiveness. In the early 2000s, J-PAL began conducting randomized evaluations to assess microcredit impacts, access, and empowerment across seven countries. It found that microcredit is not likely the best tool to improve business profitability, but it can increase freedom of choice and that more innovative credit products could lead to greater impacts on poverty (Policy Action Lab 2015).

In 1997 Mexican President Ernesto Zedillo entered office in the middle of an economic crisis. To assist the country's poorest citizens, President Zedillo and economist Santiago Levy devised PROGRESA (now known as Prospera), a program that provided cash payments to families meeting specific requirements such as keeping their children in school and visiting health clinics. The success of the program was measured through RCTs. Participants in communities who received the intervention visited health clinics at a 60 percent higher rate, with a 23 percent reduction in illness and 18 percent reduction in anemia, than those in the control group (Tollefson 2015). The RCT helped to solidify the program and now Prospera covers all of Mexico's poorest citizens. As a result of Mexico's success, the countries of Peru, Chile, Malawi, Zambia, Brazil, and Zambia have developed similar conditional cash transfer programs (Nigenda and González-Robledo 2015).

Collaborating with partners within communities to set up experiments is a good approach. Several local governments are working with academic partners or research institutions to study emerging technologies, deploy them on a small scale, gauge results, and then decide about larger-scale implementation. The Philadelphia City Planning Commission, Detroit Works Long Term Planning (recall the Detroit 24/7 participatory online game), the City of Boston, and the Cape Cod Commission partnered with Emerson College to use Community PlanIt (www.communityplanit.org), a dynamic platform that allows the public to engage in community planning. The Cape Cod Commission developed a Community PlanIt platform to understand what Cape Cod would look like in 30 years and how to make it better. Residents submitted ideas and opinions in written form that were later placed into a data visualization (www.communityplanit.org/capecod). The visualization can be distilled by gender, age, stake in the community (i.e., local employee, homeowner, student, business owner, regular visitor), and race to understand the nature of opinions.

Implementing Policy

When outputs from analysis have been interpreted, shared, and tested through experimentation, they have become actionable insights and reached their potential. At that point, it is up to decision makers to drive new, evidence-based policies and processes. Policy makers and researchers have recognized that policy making based on rigorous evidence is a better way to operate efficiently and strategically. The availability of evidence (using big data and analytics) can help planners understand what is working and what is not.

In Arizona, the City of Mesa has received numerous awards and funding for data-driven innovation related to the city's fire and medical operations (Centers for Medicare & Medicaid Services 2016). Following the recession, the city was looking for smart and effective ways to cut costs. It developed a fire and rescue response pilot program meant to treat patients with low-acuity issues (e.g., minor trauma, pain management, prescription services, immunizations) who called 911. The responders treated these patients on-site instead of transporting them to hospitals, if hospital care was not needed, and this reduced instances of high-risk patients repeatedly returning to emergency rooms. The pilot led to the Community Care Response Initiative, a program consisting of four mobile units operating 24/7, including a physician extender unit in a modified ambulance that takes teams to respond to low-acuity service calls and provide hospital discharge follow-ups. This program is powered by data collected from the emergency call-in center that allows for assessment and measurement of outcomes (Dokes 2016).

For planners, once the policy implementation process has been reached, a course of action agreeable to relevant stakeholders should have been determined. Through due diligence with analytic outputs and stakeholder engagement, the implementation of policy should begin with clear ideas about the horizon of outcomes. Table 4.4 describes characteristics of different outcome timeframes. For the long-term outcomes to happen, the short-term outcomes will need to occur and persist. This means that the policy and its evaluation will have to be spread across a sufficiently long period of time to understand if the outcomes have the desired (or undesired) impacts.

Further, policy implementation should also include constant feedback, learning, and routine adjustments. Flexibility—through continuous monitoring and policy evaluation—is required to make necessary adjustments. Especially with the ever-changing nature of data and analytics, regular monitoring and recalibration are needed. This should not be viewed as optional but as a necessary part of the operation. It is also necessary as constant monitoring can help develop insights for the next round of innovation that will occur.

Evaluating Policy Outcomes

An interesting option with big data and analytics is to continuously evaluate policies to ensure that the desired implementation and impacts are achieved. It is vital that agencies develop formal processes to evaluate policies on a regular basis. There are several reasons for this. First, conditions change regularly. Any policy that is deployed will need to be revised as conditions in a community change. Following the public protests in Ferguson, Missouri, in 2014, a US Department of Justice (DOJ) investigation of the Ferguson Police Department revealed law enforcement activities that disproportionately targeted and affected African Americans. Data on traffic stops and arrests supported many of the DOJ's findings (US Department of Justice 2015). While the civil unrest in Ferguson is an example of sudden and drastic change in the community, the subsequent investigation and analysis of data provided important insight into the law enforcement practices and policies that needed to be reevaluated and changed.

Second, feedback collected from the evaluation of policies provides opportunities to make small modifications and tweaks. This is much more advisable then waiting until it is too late, before it is necessary to overhaul the entire infrastructure and system. Consider the case of a printer. Long before a printer stops printing, it provides data through signals that the ink levels are getting lower. It is quite advis-

TABLE 4.4. OUTCOME TIMEFRAMES

Outcome Phase	Description	Trend
Short-term outcome	Changes in attitudes, knowledge, and skills	Usually tied to an intervention during the intervention (i.e., project evaluation)
Intermediate outcome	Changes in decision making	Traditionally conducted after the intervention has ended and includes more about what was learned
Long-term outcome	Behavioral and lifestyle changes	Usually measured a year or several years following the completion of an intervention

Source: Kevin Desouza and Kendra Smith

able to begin the task of getting a new cartridge well in advance of the printer running out and not being able to print. Similarly, the growth of big data, through Internet of Things technology with sensors embedded in devices, means that making these changes and modifications would be timelier when they are needed.

Currently, city infrastructure is costly and labor intensive, and infrastructure is crumbling in many American cities. For instance, the American Society for Civil Engineers gave the US drinking water infrastructure a "D" grade in its 2013 Report Card for America's Infrastructure (ASCE 2013). Currently, nearly 50 percent of the world's water disappears from leaky or aging pipes, at an annual cost of $14 billion (Goulding, Wilshere, and Starsiak n.d.). Smart-water applications might potentially be placed in water distribution and fire protection systems (like fire hydrants) to collect data on water temperature and pressure, detect leaks, and transmit that information to a smart-utility dashboard. For instance, 100 smart meters have been installed in 28 Los Angeles parks to monitor water use in real time and detect leaks more quickly by alerting city officials to performance anomalies. The investment in smart meters is an improvement from the previous method of detecting leaks by monitoring utility bills (Los Angeles 2016). If implemented, the outcomes of the analysis and implementation of policies are likely to affect water quality; aid in more efficient use of water in agriculture, which consumes 90 percent of available water; and decrease maintenance costs needed for major repairs because minor repairs can be made.

Third, evaluation is a chance to catch bad policies and decisions much more quickly. The City of Boston introduced Street Bump (www.streetbump.org), an app that detects potholes through sensors on smartphones; all users need to do is download the app and it will sense the "bumps" on its own. The app was effective in detecting potholes so crews could be sent to repair streets in a timely manner. However, there was an unintended consequence: the app inadvertently sent crews to wealthier neighborhoods because residents in these areas were more likely to use smartphones, which meant that lower-income and elderly residents were effectively nonparticipants (Luque-Ayala and Marvin 2015).

Because policy is a critical part of the big data and analytical process, it is necessary to ensure that the insights developed and that eventually mature into policies are implemented and have the policymakers' intended impact. The following are the three main types of evaluation:

1. **Policy content:** Focus on whether content clearly articulates the goals of the policy, implementation plans, and undergirding logic of how the policy will produce change.
2. **Policy implementation:** Focus on whether the policy was implemented as intended.
3. **Policy impact:** Focus on measuring whether the policy produced the intended outcomes and impact.

At any given time, one or more of these types of evaluation need to be conducted. Each type plays into the others, and their relationship is circular in nature (CDC n.d.). For instance, understanding implementation is critical to understanding effectiveness and impact. In evaluating implementation, a better understanding is necessary of why a policy did or did not have an impact and what the barriers and obstacles might have been in the implementation process.

CONCLUSION

This chapter outlined a framework for managing big data. It is important to stress that it is necessary to take great care when conducting activities in the various phases. Both individuals and organizations should increase the rigor and agility with which activities within each phase are conducted over time. Doing so will call for a deliberate effort to collect information on the time, effort, and outcomes of each activity over the course of the project in order to track improvements. Benchmarking efforts are also valuable to gauge performance within a given industry. Tracking the progress in each phase will also help in identifying areas that need investment for improvement—as you are only as good as your weakest link.

CHAPTER 5

THE FUTURE OF BIG DATA AND PLANNING

This report has discussed shifts within urban planning that make it primed for planners to try out new and promising technologies. It also outlined the fundamentals of big data, predictive analytics, Internet of Things technologies, and smart cities to illustrate the promises and opportunities inherent in these capabilities. Through tools such as modeling, simulation, machine learning, and gaming, planners can use big data to stimulate the smart-city capabilities in their communities. This report also presented a four-phase actionable framework to guide planners through the process of leveraging big data through analytics: (1) management of data sources, (2) information through analytics, (3) interpretation of the information gathered, and (4) design and implementation of policy based on analytics.

This framework is purposely generalizable and can be applied at multiple levels—from projects conducted by individuals using data within a specific domain (e.g., transportation) to larger undertakings conducted when several organizations collaborate and aggregate data and analytical capacities across multiple environments. The report also surveyed the most common analytical approaches one might deploy to generate insights from large datasets. A word of caution is in order here. The discussion of these tools was somewhat simplified so as not to make the text incomprehensible to a nontechnical reader. Readers should consult other sources, such as specific publications or online courses, to learn the nuances associated with each method before deploying any live, so as to avoid costly errors.

The future of analytics and big data is bright and will transform the practice of planning. While it is always challenging to predict with pinpoint precision how things might play out, there are several trends that warrant attention. First, there will be a growing movement to integrate traditionally disparate databases together at the city, state, and regional levels. Today, gaps exist that prevent us from getting a more holistic view of how individuals engage with public services and agencies. For example, data on car registrations might be linked to drivers' licenses but not linked with records from the department of revenue. This missing link is troublesome for several reasons, not the least of which is that it prevents cities from delivering efficient services (e.g., if someone has not paid taxes, should the state allow registering of a new car?). The linking of datasets at the local and regional levels also allows delivery of more personalized services and reflects an understanding of the dynamics of policies and regulations at a finer granularity. Simply put, layering of multiple datasets allows deeper and more precise insights into phenomena.

Second, the trend to combine analytics with know-how in terms of behavioral sciences is picking up momentum. Early in this report, this trend was alluded to when discussing the Behavioural Insights Team in the United Kingdom. Recently, in the United States, President Obama signed an executive order to promote the use of behavioral insights within the federal government to better serve the American people (Obama 2015). A key way these insights can improve government functions is through lower policy and program costs from enhanced designs that are guided by better knowledge of human behavior to help workers find better jobs, improve health outcomes, and enhance environmental sustainability.

The use of frameworks, theories, and models from behavioral science extends the benefits of analytics. Through the use of big data and predictive analytics, information on the behavior of groups becomes more precise, which will aid the efficacy of nudging interventions (Desouza and Smith 2016). Key features of nudging are that it seeks to alter individuals' behavior in predictable ways without coercion or by prohibiting any options or providing compensation. For instance, organ donation has long been a decision for individuals getting drivers' licenses where they have to opt in to participate. Nudging scholars call for a simple nudge to get more individuals to donate by requiring them to opt out rather than opt in, blurring the line of nudging and coercion.

On the one hand, a mandatory opt-in is coercive. On the other hand, the opportunity to opt out is considered choice architecture, or the presentation of choices in different ways to achieve different outcomes (Thaler and Sunstein 2008). In the future, the fine line between nudging and coercion and how to address complex social and ethical issues related to these concepts will be a critical focus for many researchers and policy makers.

Data and insights will provide the ability to spot early signs of trouble or things going astray (e.g., a student who might be on a course to drop out of college or a recently released prisoner about to commit another crime). The question becomes how and when to intervene and how severely to do so. Glimpses of this tension have emerged in more recent events, such as the battle between Apple and the Federal Bureau of Investigation (FBI) over unlocking the phone of San Bernardino mass shooter Syed Rizwan Farook (New York Times 2016). Apple's pushback against the request is emblematic of greater issues about data capabilities and use of those capabilities, which will likely be debated in the future.

Third, the trend of the automation of almost everything is in full swing. For example, in Los Angeles, the Automated Traffic Surveillance and Control system (ATSAC) is located in the downtown areas of the city (Bliss 2014). ATSAC monitors and adjusts the timing of traffic signals at intersections, which reduces delays. The streets are outfitted with sensors that detect the movement of any vehicle. The system has 450 closed-circuit cameras that oversee the city's intersections and loop detectors that pick up information about the number of vehicles passing, traffic speed, and congestion levels. In addition to this, underground fiber optics link all of the traffic signals as well as send data to city engineers.

Systems like ATSAC are just the beginning of what is expected to be a serious move to automation in the future. Autonomous vehicles already have the potential to change the fabric of cities and communities not only directly (e.g., how people move) but indirectly as well. Consider the disruptions that will occur as revenue bases of cities change due to loss of revenue from sources such as speeding ticket fines and driver's license and car registration fees when fewer people need to own vehicles and instead use evolving shared modes for travel. These technological disruptions will also affect entire industries, such the auto insurance industry.

Automation will also affect how data are collected and who analyzes the data. Today, most of the data analysis that occurs in the context of planning with big data is to some degree a combination of human and machine intelligence. In the financial sector, as a contrast, the analysis is more driven by machines than humans for most of the traditional large-scale data applications (this is due to the fact that this sector has a longer history of dealing with large-scale data). One trend to watch out for is the infusing of more machine-driven analytics that guide behavior or automated devices. Think back to the 2004 movie *I, Robot*, which took place in 2035, where robots were built to protect and serve humanity. Instead, they ended up making life and death decisions based on the analytics coded into them. Automation offers great promise if done properly, but there will be factors such as the changing nature of what planners do and the need to ensure that they have knowledge and skills to apply beyond what machines can do.

Fourth, given how much people's lives, organizations, and societies have come to depend on the digital information infrastructure, it is not surprising that the unscrupulous want to disrupt and cause harm to these systems. Hacks across the globe in both the public and private sectors have proven that this harm is clear and present. Actions taken by leaders to develop taskforces to combat the hacking and misuse of data are steps in the right direction, but more and more evidence suggests that the "bad people" are winning. This means that, along with new technologies, it is necessary to upgrade our information technology infrastructure to match our needs and the times in which we live. Further, given that most technology and systems that we use are designed by private-sector actors (e.g., IBM, Apple, Google), partnerships and collaborative governance approaches to safety and security are paramount. The flare up between Apple and the FBI is likely to be one of many in the near future. In the future, municipalities and different government agencies will need to find ways to coexist; for instance, officials concerned with big data and other technologies must learn to function effectively with officials concerned with privacy and security.

Fifth, the rise of the nontraditional planner and the growth of crowdsourcing platforms mean that we are likely to see many more solutions developed outside of local government. Although this might be a tad intimidating, this is a good thing. On the one hand, more people are caring about their local communities and using their talent and knowledge to leverage data and solve challenges. On the other hand, if planners rebuff these actions and contributions, the result will likely be significant clashes because citizens are not likely to back down from their roles as the new "powerbrokers" in the community. Planners must change their orientation to issues and closed-door policy making. They must instead become more engaged with the community as a valuable resource to them.

Urban planners play important roles in most cities as they create and implement visions for communities. They lead efforts around social change, policy making, community empowerment, sustainable growth, innovation, disaster preparedness, economic development, and resilience. Planners also solve problems through research, development, and design. Fundamental to carrying out these activities is their ability to make decisions in an effective and efficient manner. To increase their decision-making capacities, it is critical for planners to leverage data in innovative ways. The future offers exciting possibilities, and planners are encouraged to stay abreast of developments in data science and to find ways to begin conversations with city leadership about how these technology-based opportunities can advance their communities.

REFERENCES

Accenture. 2013. "Accenture Survey 2013: What Travelers Want from Public Transport Providers." Available at www.accenture.com/us-en/insight-acn-public-transport-survey.aspx.

———. 2014. *Insurance Telematics: A Game-Changing Opportunity for the Industry*. Available at http://ins.accenture.com/rs/accenturefs/images/Insurance-telematics-A%20game-changing-opportunity-for-the-industry.pdf.

Acquisti, Alessandro, and Ralph Gross. 2009. "Predicting Social Security Numbers from Public Data." *PNAS* 106 (27): 10975–10980.

ACSE (American Association of Civil Engineers). 2013. "Wastewater: 2103 Grade D." Available at www.infrastructurereportcard.org/a/#p/wastewater/overview.

Adshead, Antony. 2014. "Data Set to Grow 10-Fold by 2020 as Internet of Things Takes Off." *Computer Weekly*, April 9. Available at www.computerweekly.com/news/2240217788/Data-set-to-grow-10-fold-by-2020-as-internet-of-things-takes-off.

Agard, Bruno, Catherine Morency, and Martin Trépanier. 2007. *Mining Public Transport User Behaviour from Smart Card Data*. Montreal, Quebec: CIRRELT. Available at www.cirrelt.ca/DocumentsTravail/CIRRELT-2007-42.pdf.

Ahmed, Murad. 2015. "Mixed Signals on London's Drive for Smart City Technology." *Financial Times*, October 14. Available at www.ft.com/cms/s/2/8fb15e14-6604-11e5-a57f-21b88f7d973f.html#axzz3w6OdeV00.

Alloway, Tracy. 2015. "Big Data: Credit Where Credit's Due." *Financial Times*, February 4. Available at https://next.ft.com/content/7933792e-a2e6-11e4-9c06-00144feab7de.

An, Li, Marc Linderman, Ashton Shortridge, Jiaguo Qi, and Jianguo "Jack" Liu. 2005. "Exploring Complexity in a Human-Environment System: An Agent-Based Spatial Model for Multidisciplinary and Multiscale Integration." *Annals of the Association of American Geographers* 95 (1): 54–79.

Anderson, Chris. 2008. "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete." *Wired*, June 23. Available at http://www.wired.com/2008/06/pb-theory/.

Anderson, Monica. 2015. *Technology Device Ownership: 2015*. Washington, DC: Pew Research Center. Available at www.pewinternet.org/2015/10/29/technology-device-ownership-2015.

Araya, Daniel. 2015. "Interview: Tim O'Reilly Talks Algorithmic Regulation, Cyber-Terrorism, and Why He Doesn't Like the Term 'Automation.'" *Futurism*, September 14. Available at http://futurism.com/interview-tim-oreilly-talks-algorithmic-regulation-cyber-terrorism-and-why-he-doesnt-like-the-term-automation.

Arnett, George. 2014. "Glasgow Has Lowest Life Expectancy in the UK." *Guardian*, April 16. Available at www.theguardian.com/news/datablog/2014/apr/16/commonwealth-games-2014-glasgow-lowest-life-expectancy-uk.

Ash, Lucy. 2014. "Why Is Glasgow the UK's Sickest City?" *BBC News Magazine*, June 5. Available at www.bbc.com/news/magazine-27309446.

Badger, Emily. 2012. "How to Catch a Criminal with Data." *CityLab*, March 14. Available at www.citylab.com/tech/2012/03/how-catch-criminal-data/1477/.

Batty, Michael. 2012. "Building a Science of Cities." *Cities*, 29 (1): S9–S16.

BBC News. 2015. "London's Population Hits 8.6M Record High." *BBC News*, February 2. Available at www.bbc.com/news/uk-england-london 31082941.

Best, Jo. 2016. "Who Really Owns Your Internet of Things Data?" *ZDNet*, January 11. Available at www.zdnet.com/article/who-really-owns-your-internet-of-things-data/

Bettencourt, Luis. 2013. "The Uses of Big Data in Cities." Working Paper 2013-09-029, Santa Fe Institute, Santa Fe, NM. Available at www.santafe.edu/media/workingpapers/13-09-029.pdf.

Biermann, Frank. 2007. "Earth System Governance's a Crosscutting Theme of Global Change Research." *Global Environmental Change* 17 (3): 326–337.

Biewald, Lukas. 2015. "The Data Science Ecosystem Part 2: Data Wrangling." *ComputerWorld*, April 1. Available at www.computerworld.com/article/2902920/the-data-science-ecosystem-part-2-data-wrangling.html.

Bliss, Laura. 2014. "LA's Automated Traffic Surveillance and Control System." *Los Angeles Magazine*, May 21. Available at www.lamag.com/citythinkblog/crossed-signals.

Block, Dustin. 2012. "Detroit 24/7 Online Game Signs Up 550 People; Mission No. 2 Starts Monday." *MLive*, May 12. Available at www.mlive.com/business/detroit/index.ssf/2012/05/detroit_247_online_game_signs.html.

Blumenstock, Joshua, Gabriel Cadamuro, and Robert On. 2015. "Predicting Poverty and Wealth from Mobile Phone Metadata." *Science* 350 (6264), 1073–1076.

boyd, danah, and Kate Crawford. 2012. "Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon." *Information, Communication & Society* 15 (5): 662–679.

Brookings (Brookings Institution Center on Urban and Metropolitan Policy). 2000. *Moving Beyond Sprawl: The Challenge for Metropolitan Atlanta*. Washington, DC: Brookings Institution. Available at www.brookings.edu/~/media/research/files/reports/2000/3/atlanta/atlanta.pdf.

Castro, Daniel. 2014. "Comments in Response to the Office of Science and Technology Policy's (OSTP) Request for Public Comment on the Public Policy Implications of 'Big Data.'" Available at www2.datainnovation.org/2014-ostp-big-data-cdi.pdf.

CDC (Centers for Disease for Control and Prevention). n.d. *Step by Step—Evaluating Violence and Injury Prevention Policies*. Available at www.cdc.gov/injury/pdfs/policy/Brief%201-a.pdf.

Centers for Medicare & Medicaid Services. 2016. "Health Care Innovation Awards Round Two: Arizona." Available at https://innovation.cms.gov/initiatives/Health-Care-Innovation-Awards-Round-Two/Arizona.html.

Choi, Charles. 2015. "Mobile Phone Data Predicts Poverty in Rwanda." *IEEE Spectrum*, November 30. Available at http://spectrum.ieee.org/tech-talk/consumer-electronics/portable-devices/mobile-phone-data-predicts-poverty-in-rwanda.

Cisco. 2015. *The Zettabyte Era—Trends and Analysis*. Available at www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.html.

City Form Lab. 2016. "Urban Network Analysis Toolbox for ArcGIS." Available at http://cityform.mit.edu/projects/urban-network-analysis.html.

Clarke, Ruthbea Yesner, Yayoi Hirose, Gerald Wang, Baogui Ding, Lianfeng Wu, and Aparna Sree. 2014. *IDC FutureScape: Worldwide Smart Cities 2015 Predictions*. Farmingham, MA: IDC Research.

Copenhagen, City of (Denmark). 2016. "Livable Green City." Available at http://international.kk.dk/artikel/livable-green-city.

Crawford, Kate, and Megan Finn. 2015. "The Limits of Crisis Data: Analytical and Ethical Challenges of Using Social and Mobile Data to Understand Disasters." *GeoJournal* 80 (4): 491–502.

Creswell, John W., and Vicki L. Plano Clark. 2011. *Designing and Conducting Mixed Methods Research*. Thousand Oaks, CA: Sage Publications.

Crossrail. 2016. "Driving Industry Standards for Design Innovation on Major Infrastructure Projects." Available at www.crossrail.co.uk/construction/building-information-modelling.

Cukier, Kenneth Neil, and Viktor Mayer-Schoenberger. 2013. "The Rise of Big Data." *Foreign Affairs*, May/June. Available at www.foreignaffairs.com/articles/2013-04-03/rise-big-data.

Data Kind. 2012. "Out On a Limb—For Data." Available at www.datakind.org/projects/out-on-a-limb-for-data.

DeAngelis, Stephen. 2014. "Big Data Is Essential for Improving Urban Efficiency." *Enterra Solutions* (blog), September 23. Available at www.enterrasolutions.com/2014/09/big-data-essential-improving-urban-efficiency.html.

Desai, Anand, and James E. Storbeck. 1990. "A Data Envelopment Analysis for Spatial Efficiency." *Computers, Environment and Urban Systems* 14 (2): 145-156.

Desouza, Kevin C. 2006. "Knowledge Management Maturity Model: Theoretical Development and Preliminary Empirical Testing." Thesis, PhD in Management Information Systems, University of Illinois at Chicago.

Desouza, Kevin C., and Kendra L. Smith. 2016. "Predictive Analytics: Nudging, Shoving, and Smacking Behaviors in Higher Education." *Educause Review*, August 22. Available at http://er.educause.edu/articles/2016/8/predictive-analytics-nudging-shoving-and-smacking-behaviors-in-higher-education?zbrandid=4007&zidType=CH&zid=37313605&zsubscriberId=1008492045&zbdom=http://educause.informz.net.

Desouza, Kevin C., and Paul Simons. 2014. *Mobile App Development in Highly Regulated Industries: Risks, Rewards and Recipes*. Mount Laurel, NJ: Society for Information Management.

Desouza, Kevin, David Swindell, Kendra Smith, Alison Sutherland, Kena Fedorschak, and Carolina Coronel. 2015. *Local Government 2035: Strategic Trends and Implications of New Technologies*. Washington, DC: Brookings Institution. Available at www.brookings.edu/~/media/research/files/papers/2015/05/29-local-government-strategic-trends-desouza/desouza.pdf.

Deutscher, Maria. 2014. "IoT Will Propel Digital Universe Past 40ZB by 2020." *Silicon Angle*, April 15. Available at http://siliconangle.com/blog/2014/04/15/iot-will-propel-digital-universe-past-40zb-by-2020.

Dier, Arden. 2016. "Woman's Fitbit Reveals She's Pregnant." *Yahoo News*, February 9. Available at www.yahoo.com/health/womans-fitbit-reveals-shes-pregnant-151659615.html.

Dokes, Jennifer. 2016. *Fired Up: Community Paramedicine Models Blaze a Trail for Healthcare Delivery Reform*. Available at www.naemt.org/docs/default-source/community-paramedicine/Toolkit/st-lukes-health-initiative---community-paramedicine-february-2016-pd.pdf?status=Temp&sfvrsn=0.04255103918986358.

Dutcher, Jennifer. 2014. "What is Big Data?" *datascience@berkeley* (blog), September 3. Available at https://datascience.berkeley.edu/what-is-big-data.

Eagleburger, Amy. 2015. "DC Tackles Bus Bunching with Smartphones." *GCN*, September 21. Available at https://gcn.com/articles/2015/09/21/bus-bunching.aspx.

Esri. 2011. "Open Data." Available at www.esri.com/~/media/files/pdfs/library/brochures/pdfs/open-data-gov20.pdf.

Falk, Tyler. 2013. "Was South Korea's Smart City Experiment a Success?" *ZDNet,* September 5. Available at www.zdnet.com/article/was-south-koreas-smart-city-experiment-a-success.

Feuer, Alan. 2013. "The Mayor's Geek Squad." *New York Times*, March 23. Available at www.nytimes.com/2013/03/24/nyregion/mayor-bloombergs-geek-squad.html?pagewanted=all.

Figliozzi, Miguel A., and Bryan Philip Blanc. 2015. *Evaluating the Use of Crowdsourcing as a Data Collection Method for Bicycle Performance Measures and Identification of Facility Improvement Needs*. Washington DC: US Department of Transportation.

Fisher, H. B., and Gerard Rushton. 1979. "Spatial Efficiency of Service Locations and the Regional Development Process." *Papers in Regional Science* 42 (1): 83-97.

Galaxy Zoo. n.d. "The Story So Far." Available at www.galaxyzoo.org/#/story.

Gardiner, Bonnie. 2015. "NSW to See Australia's First Government Data Analytics Centre." *CIO*, August 4. Available at www.cio.com.au/article/581132/nsw-see-australia-first-government-data-analytics-centre.

Gartner. 2012. "Gartner Says by 2014, 80 Percent of Current Gamified Applications Will Fail to Meet Business Objectives Primarily Due to Poor Design." *Gartner*, November 27. Available at www.gartner.com/newsroom/id/2251015.

———. 2014. "Gartner Says the Internet of Things Will Transform the Data Center." *Gartner Newsroom*, March 19. Available at www.gartner.com/newsroom/id/2684616.

GCN Staff. 2015. "Following Attack, Intruders Still Cooling in State Department Network." *GCN*, February 23. https://gcn.com/articles/2015/02/23/state-dept-hack.aspx.

George, John. 2016. "Philadelphia Union Teams Up with Uber." *Philadelphia Business Journal*, March 14. Available at www.bizjournals.com/philadelphia/news/2016/03/14/philadelphia-union-uber-rideshare-mls-soccer-taxi.html.

Gibbs, Samuel. 2016. "Dropbox Hack Leads to Leaking of 68M User Passwords on the Internet." *Guardian*, August 1. Available at www.theguardian.com/technology/2016/aug/31/dropbox-hack-passwords-68m-data-breach.

Glaeser, Edward L., Scott Duke Kominers, Michael Luca, and Nikhil Naik. 2015. "Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life." Working Paper 16-065, Harvard Business School, Cambridge, MA. Available at http://people.hbs.edu/mluca/BigDataBigCities.pdf.

Glasgow Centre for Population Health. 2016. "Scottish 'Excess' Mortality: Comparing Glasgow with Liverpool and Manchester." Available at www.gcph.co.uk/work_themes/theme_1_understanding_glasgows_health/excess_mortality_comparing_glasgow.

Glinton, Sonari. 2015. "How a Little Lab in West Virginia Caught Volkswagen's Big Cheat." *NPR*, September 24. Available at www.npr.org/2015/09/24/443053672/how-a-little-lab-in-west-virginia-caught-volkswagens-big-cheat.

Global Construction Review. 2014. "Researchers Unveil Drones that Minutely Inspect High-Rise Buildings." *Global Construction Review*, July 18. Available at www.globalconstructionreview.com/innovation/researchers-unveil-drones-minutely-inspect353678.

Glum, Julia. 2015. "Millennials Selfies: Young Adults Will Take More Than 25,000 Pictures of Themselves During Their Lifetimes: Report." *International Business Times*, September 22. Available at www.ibtimes.com/millennials-selfies-young-adults-will-take-more-25000-pictures-themselves-during-2108417.

Goddin, Paul. 2014. "Wearable Tech Will Make Our Cities More Walkable." *Mobility Lab Express*, November 3. Available at http://mobilitylab.org/2014/11/03/wearable-tech-will-make-our-cities-more-walkable.

Goulding, Charles, Michael Wilshere, and Adam Starsiak. n.d. "The R&D Tax Credit Aspects of Water Analytics." *R&D Taxsavers* (blog). Available at www.rdtaxsavers.com/articles/Water-Analytics.

GRAPHISOFT. 2015. "GRAPHISOFT Hosts Live Online Webinar: Energizing Heritage Conservation with ARCHICAD Point Clouds." Available at www.graphisoft.com/info/news/press_releases/graphisoft-hosts-live-online-webinar-energizing-heritage-conservation-with-archicad-point-clouds-and-bim.html.

Greater London Authority. 2015. *Smart London Projects*. Available at www.london.gov.uk/sites/default/files/add348_smart_london_projects_signed_pdf.pdf.

———. 2016. "Energy Supply." Available at www.london.gov.uk/what-we-do/environment/energy/energy-supply.

Greater New Orleans Fair Housing Action Center. 2016. *Practical Steps to End Poverty for Families in the Housing Choice Voucher Program*. Available at www.gnofairhousing.org/wp-content/uploads/2016/01/HANO-HCVP_Issue_Brief_Jan-2016.pdf.

Greenberg, Andy. 2015. "Hackers Remotely Kill a Jeep on the Highway—With Me in It." *Wired*, July 21. Available at www.wired.com/2015/07/hackers-remotely-kill-jeep-highway.

Grubesic, Tony, Ran Wei, Alan Murray, and Fangwu Wei. 2016. "Essential Air Service in the United States: Exploring Strategies to Enhance Spatial and Operational Efficiencies." *International Regional Science Review* 39 (1): 108–130.

Hellerstein, Joseph M. 2008. "Quantitative Data Cleaning for Large Databases." Unpublished paper, EESC Computer Science Division, UC Berkeley. Available at http://db.cs.berkeley.edu/jmh/papers/cleaning-unece.pdf.

Hill, Kashmir. 2011. "Fitbit Moves Quickly After Users' Sex Stats Exposed." *Forbes*, July 5. Available at www.forbes.com/sites/kashmirhill/2011/07/05/fitbit-moves-quickly-after-users-sex-stats-exposed/#216949b879e7.

Hitachi. 2015. "Hitachi Data Systems Unveils New Advancements in Predictive Policing to Support Safer, Smarter Societies." Available at www.hds.com/en-us/news-insights/press-releases/2015/gl150928.html.

Horrigan, John B., and Maeve Duggan. 2015. *Home Broadband 2015: The Share of Americans with Broadband at Home Has Plateaued, and More Rely Only on Their Smartphones for Online Access*. Washington, DC: Pew Research Center. Available at www.pewinternet.org/2015/12/21/home-broadband-2015.

Horst, Peter, and Robert Duboff. 2015. "Don't Let Big Data Bury Your Brand." *Harvard Business Review*, November. Available at https://hbr.org/2015/11/dont-let-big-data-bury-your-brand.

Huang, Qingxu, Dawn C. Parker, Shipeng Sun, and Tatiana Filatova. 2013. "Effects of Agent Heterogeneity in the Presence of a Land-Market: A Systematic Test in an Agent-Based Laboratory." *Computers, Environment, and Urban Systems* 41 (September): 188–203.

Iannucci, Corrado, Luca Congedo, and Michele Munafò. 2012. "Urban Sprawl Indicators and Spatial Planning: The Data Interoperability in INSPIRE and Plan4all." In *Planning Support Tools: Policy Analysis, Implementation and Evaluation*, Proceedings of the Seventh International Conference on Informatics and Urban and Regional Planning INPUT2012, 583–594. University of Cagliari, Sardina, Italy, May 10–12.

IBM. 2016. "IBM SPSS Modeler: Product Pricing." Available at www.ibm.com/marketplace/cloud/spss-modeler/purchase/us/en-us#product-header.

IDC (Internatonal Data Corporation). 2014. "IDC Reveals Worldwide Internet of Things Predictions for 2015." Business Wire, December 3. Available at http://www.businesswire.com/news/home/20141203006197/en/IDC-Reveals-Worldwide-Internet-Predictions-2015.

Ingis, Stuart, Michael A. Signorelli, and Ariel S. Wolf. 2015. "Vehicle Privacy: FTC Weighs in on V2V Technology and Sen. Schumer Introduces GPS Legislation" *Lexology*, January 6. Available at www.lexology.com/library/detail.aspx?g=cbba7298-6312-4742-98a5-9fe484090797.

Insurance Journal. 2015. "Insurers See Impact from Wearable Devices within 2 Years." *Insurance Journal*, May 6. Available at www.insurancejournal.com/news/national/2015/05/06/367020.htm.

International City/County Management Association. 2016. "Technology." Available at http://icma.org/en/icma/knowledge_network/topics/kn/Topic/137/Technology.

Internet World Stats. 2016. "Top 50 Countries with the Highest Internet Penetration Rates—2013." Available at www.internetworldstats.com/top25.htm.

Jacobs, Jane. 1961. *Death and Life of Great American Cities*. New York: Random House.

Jin, Xiongbing. 2010. "Modelling the Influence of Neighbourhood Design on Daily Trip Patterns in Urban Neighbourhoods." PhD diss., Memorial University of Newfoundland. Available at ftp://ftp.ci.austin.tx.us/Walkability_PDR/Jin_NeighbDesign_TripPattern.pdf.

Kharpal, Arjun. 2015. "Year of the Hack? A Billion Records Compromised in 2014." *CNBC*, February 12. Available at www.cnbc.com/2015/02/12/year-of-the-hack-a-billion-records-compromised-in-2014.html.

King, Hope. 2016. "Woman's Home Demolished after Google Maps Error." *CNNMoney*, March 25. Available at http://money.cnn.com/2016/03/25/technology/google-maps-house/.

Kirchner, Lauren. 2015. "When Discrimination Is Baked into Algorithms." *The Atlantic*, September 6. Available at www.theatlantic.com/business/archive/2015/09/discrimination-algorithms-disparate-impact/403969.

Knight, Will, 2015. "This Car Knows Your Next Misstep Before You Make It." *MIT Technology Review*, October 1. Available at www.technologyreview.com/s/541866/this-car-knows-your-next-misstep-before-you-make-it.

Konsynski, Benn R., and F. Warren McFarlan. 1989. "Information Partnerships—Shared Data, Shared Scale." *Harvard Business Review*, 68 (5): 114–120.

Kulkarni, Siddharth, and Kenan Fikri. 2015. *GDP Contractions Aside, Recovery Continues in Most Metro Areas*. Washington, DC: Brookings Institution. Available at www.brookings.edu/blogs/the-avenue/posts/2015/07/20-gdp-metro-area-recovery-kulkarni-fikri.

Kurkjian, Katie M., Michelle Winz, Jun Yang, Kate Corvese, Ana Colón, Seth J. Levine, Jessica Mullen, Donna Ruth, Rexford Anson-Dwamena, Tesfaye Bayleyegn, and David S Chang. 2016. "Assessing Emergency Preparedness and Response Capacity Using Community Assessment for Public Health Emergency Response Methodology: Portsmouth, Virginia, 2013." *Disaster Medicine and Public Health Preparedness* 10 (20): 193–198.

Kuwait-MIT Center for Natural Resources and the Environment. 2015. "Real-Time Epidemiology from Urban Wastewater." *MIT News*, November 2. Available at http://news.mit.edu/2015/real-time-urban-epidemiology-from-wastewater-1102.

Langley, Ann, Henry Mintzberg, Patricia Pitcher, Elizabeth Posada, and Jan Saint-Macary. 1995. "Opening Up Decision Making: The View from the Black Stool." *Organization Science* 6 (3): 260–279.

Lazer, David, and Ryan Kennedy. 2015 "What We Can Learn from the Epic Failure of Google Flu Trends." *Science*, October 1. Available at www.wired.com/2015/10/can-learn-epic-failure-google-flu-trends.

Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. "The Parable of Google Flu: Traps in Big Data Analysis." *Science* 343 (March 14). Available at http://gking.harvard.edu/files/gking/files/0314policyforumff.pdf?m=1394735706.

Leyden, John. 2016. "Smart City Transport Infrastructure." *Register*, April 22. Available at www.theregister.co.uk/2016/04/22/smart_transport_hackable.

Lohr, Steve. 2014. "For Big-Data Scientists, 'Janitor Work' Is Key Hurdle to Insights." *New York Times*, August 17. Available at www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html.

Lorenz, Jan, Heiko Rauhut, Frank Schweitzer, and Dirk Helbing. 2011. "How Social Influence Can Undermine the Wisdom of Crowd Effect." *PNAS* 108 (22): 9020–9025.

LosAngeles,Cityof(California).2016."MayorGarcettiAnnouncesInnovative Pilot Program to Conserve More Water in Historic Drought." Available at https://www.lamayor.org/mayor-garcetti-announces-innovative-pilot-program-conserve-more-water-historic-drought.

Luque-Ayala, Andres, and Simon Marvin. 2015. "Developing a Critical Understanding of Smart Urbanism?" Urban Studies 52 (12): 2105–2116.

Macdonell, Hamish. 2015, "Glasgow: The Making of a Smart City." *Guardian*, April 21. Available at www.theguardian.com/public-leaders-network/2015/apr/21/glasgow-the-making-of-a-smart-city.

Maddox, Teena. 2015. "The World's Smartest Cities: What IoT and Smart Governments Will Mean for You." *Tech Republic*, November 10. Available at www.techrepublic.com/article/smart-cities.

Manyika, James, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela H. Byers. 2011. *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. New York: McKinsey & Co. Available at www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.

Manyika, James, Michael Chui, Peter Bisson, Jonathan Woetzel, Richard Dobbs, Jacques Bughin, and Dan Aharon. 2015. *Unlocking the Potential of the Internet of Things*. New York: McKinsey & Company. Available at www.mckinsey.com/business-functions/digital-mckinsey/our-insights/the-internet-of-things-the-value-of-digitizing-the-physical-world.

Maroulis, Spiro, Eytan Bakshy, Louis Gomez, and Uri Wilensky. 2014. "Modeling the Transition to Public School Choice." *Journal of Artificial Societies and Social Simulation* 17 (2).

Maroulis, Spiro, Roger Guimera, Heywood Petry, Luis Gomez, Luis A. N. Amaral, and Uri Wilensky. 2010. "A Complex Systems View of Educational Policy." *Science* 330 (6000): 38–39.

Mastercard. 2015. "Cubic and MasterCard Launch the Urbanomics Mobility Project." *Engagement Bureau*, September 15. Available at http://newsroom.mastercard.com/press-releases/cubic-and-mastercard-launch-the-urbanomics-mobility-project.

Maus, Jonathan. 2014. "ODOT Embarks on 'Big Data' Project with Purchase of Strava Dataset." *Bikeportland.org* (blog), May 1. Available at http://bikeportland.org/2014/05/01/odot-embarks-on-big-data-project-with-purchase-of-strava-dataset-105375.

Maxwell, Rebecca. 2015. "Big Data, GIS, and Bikes." *GIS Lounge*, April 22. Available at www.gislounge.com/using-big-data-and-gis-to-plan-bike-routes.

McCarthy, James. 2005. *Traffic Analysis Toolbox*. Washington, DC: US Department of Transportation. Available at www.fhwa.dot.gov/publications/publicroads/05mar/08.cfm.

McElvaney, Shannon.2012. *Geodesign: Case Studies in Regional and Urban Planning*. Redlands, CA: Esri Press.

Metz. Rachel. 2015. "The Struggle for Accurate Measurements on Your Wrist." *MIT Technology Review*, June 22. Available at www.technologyreview.com/s/538416/the-struggle-for-accurate-measurements-on-your-wrist.

Moraff, Christopher. 2014. "The Problem with Some of the Most Powerful Numbers in Modern Policing." *New City*, December 15. Available at https://nextcity.org/daily/entry/predictive-policing-crime-stats-data-measure.

Morrison, Pamela D., John H. Roberts, and David F. Midgley. 2004. "The Nature of Lead Users and Measurement of Leading Edge Status." *Research Policy* 33 (2): 351–362.

Mullich, Joe. 2013. "Closing the Big Data Gap in Public Sector." *Bloomberg Businessweek Research Services*, September. Available at www.sap.com/bin/sapcom/hu_hu/downloadasset.2013-09-sep-23-16.closing-the-big-data-gap-in-public-sector-pdf.html.

Mun, Min, Sasank Reddy, Katie Shilton, Nathan Yau, Jeff Burke, Deborah Estrin, Mark Hansen, Eric Howard, Ruth West, and Peter Boda. 2009. "PEIR, the Personal Development Impact Report, as a Platform for Participatory Sensing Systems Research." Paper presented at MobiSys '09, Krakow, Poland, June 22–25. Available at www.cs.cornell.edu/~destrin/resources/conferences/2009-jun-mun-sheddy-peir.pdf.

Municipal Art Society of New York. 2016. "Accidental Skyline." Available at www.mas.org/urbanplanning/accidental-skyline.

Nakashima, Ellen. 2016. "Powerful NSA Hacking Tools Have Been Revealed Online." *Washington Post*, August 16. Available at www.washingtonpost.com/world/national-security/powerful-nsa-hacking-tools-have-been-revealed-online/2016/08/16/bce4f974-63c7-11e6-96c0-37533479f3f5_story.html.

Neumann, Carl-Stefan. 2015. "Big Data Versus Big Congestion: Using Information to Improve Transport." Available at www.mckinsey.com/insights/infrastructure/big_data_versus_big_congestion_using_information_to_improve_transport.

New York Times. 2016. "Breaking Down Apple's iPhone Fight with the US Government." *New York Times*, March 21. Available at www.nytimes.com/interactive/2016/03/03/technology/apple-iphone-fbi-fight-explained.html.

NHTSA (National Highway Traffic Safety Administration). n.d. *Preliminary Statement of Policy Concerning Automated Vehicles*. Available at http://orfe.princeton.edu/~alaink/SmartDrivingCars/Automated_Vehicles_Policy.pdf.

———. 2014. *Vehicle-to-Vehicle Communications: Readiness of V2V Technology for Application*. Available at www.safercar.gov/staticfiles/safercar/v2v/V2V_Fact_Sheet_101414_v2a.pdf.

Nigenda, Gustavo, and Luz María González-Robledo. 2015. *Lessons Offered by Latin American Cash Transfer Programmes, Mexico's Oportunidades, and Nicaragua's SPN: Implications for African Countries*. London: DFID Health Systems Resource Centre.

Obama, Barack. 2015. *Executive Order—Using Behavioral Science Insights to Better Serve the American People*. Available at www.whitehouse.gov/the-press-office/2015/09/15/executive-order-using-behavioral-science-insights-better-serve-american.

OECD. 2016. *Car Purchase Tax: Green Tax Reform in Israel*. Available at www.oecd.org/environment/tools-evaluation/OECDWorkingPaper-Green-Tax-Reform-in-Israel.pdf.

Olvasrud, Thor. 2013. "New York Turns to Big Data to Solve Big Tree Problem." *CIO*, June 4. Available at www.cio.com/article/2385245/data-management/new-york-turns-to-big-data-to-solve-big-tree-problem.html.

Pennington, Sylvia. 2014. "Can Emergencies Be Predicted? Miinder Thinks So." *Sydney Morning Herald*, March 13. Available at www.smh.com.au/it-pro/government-it/can-emergencies-be-predicted-miinder-thinks-so-20140313-hvi3s.html.

Perrin, Andrew. 2015. *Social Media Usage: 2005-2015*. Washington, DC: Pew Research Center. Available at www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015.

Policy Action Lab. 2015. *Where Credit Is Due*. Available at www.povertyactionlab.org/publication/where-credit-is-due.

Potash, Eric, Joe Brew, Alexander Loewi, Subhabrata Majumdar, Andrew Reece, Joe Walsh, Eric Rozier, Emile Jorgenson, Raed Mansour, and Rayid Ghani. 2015. "Predictive Modeling for Public Health: Preventing Childhood Lead Poisoning." In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2039–2047. Sydney, Australia, August 10–13. Available at http://dssg.uchicago.edu/wp-content/uploads/2016/01/p2039-potash.pdf.

Pultar, Edward. 2016. "Valarm Tools Cloud + Web Dashboards Using Campbell Scientific and Vaisala Industrial IoT Sensors for Remotely Monitoring Weather and Wild Fire Risk." Available at www.valarm.net/blog/valarm-tools-cloud-web-dashboards-using-campbell-scientific-and-vaisala-industrial-iot-sensors-for-remotely-monitoring-weather-and-wild-fire-risk/.

Rahm, Erhard, and Hong Hai Do. 2000. "Data Cleaning: Problems and Current Approaches." *IEEE Computer Society Bulletin of the Technical Committee on Data Engineering* 23 (4): 3–13.

Reid, Michael. 2011. "Behind the Glasgow Effect." *Bulletin of the World Health Organization* 89: 701–776. Available at www.who.int/bulletin/volumes/89/10/11-021011/en.

Rein, Lisa. 2015. "The Federal Government Wants You to Review It on Yelp." *Washington Post*, August 18. Available at www.washingtonpost.com/news/federal-eye/wp/2015/08/18/the-federal-government-wants-you-to-review-it-on-yelp.

Renn, Aaron M. 2015. "The Other Digital Divide." *Governing*, May. Available at www.governing.com/columns/eco-engines/gov-digital-divide-small-big-cities.html.

Roberts, Neil. 2013. "How Big Data Keeps Yarra Trams Running on Time, Rain or Shine." *Building a Smarter Planet* (blog), September 16. Available at http://asmarterplanet.com/blog/2013/09/yarra.html.

Robinson, Jessica. 2013. "Search for Missing Idaho Hiker Turns to Drones, Crowdsourcing." *NW News Network*, October 16. Available at http://nwnewsnetwork.org/post/search-missing-idaho-hiker-turns-drones-crowdsourcing.

Roy, Deb. 2011. "The Birth of a Word." Available at www.ted.com/talks/deb_roy_the_birth_of_a_word/transcript?language=en.

Rutkin, Aviva. 2016. "Pic-Scanning AI Estimates City Air Pollution from Mass of Photos." *New Scientist*, February. Available at www.newscientist.com/article/2076562-pic-scanning-ai-estimates-city-air-pollution-from-mass-of-photos.

SAP. 2016. "Keeping Boston Safer, Cleaner, and More Productive." Available at www.sap.com/customer-testimonials/public-sector/city-boston.html.

Schreier, Martin, Stefan Oberhauser, and Reinhard Prügl. 2007. "Lead Users and the Adoption and Diffusion of New Products: Insights from Two Extreme Sports Communities." *Marketing Letters* 18 (1–2): 15-30.

Selkirk, Diane. 2016. "The End of the Virtual Search Party?" *Men's Journal*, Collections. Available at www.mensjournal.com/expert-advice/the-end-of-the-virtual-search-party-20140327.

SFCTA (San Francisco County Transportation Authority). 2016. "CycleTracks for iPhone and Android." Available at www.sfcta.org/modeling-and-travel-forecasting/cycletracks-iphone-and-android.

Shanteau, James. 1988. "Psychological Characteristics and Strategies of Expert Decision Makers." *Acta Psychologica* 68 (1–3): 203–215.

Silda, Oliver. 2009. "Improved Pedestrian Tracking for Urban Planning." *SPIE*, December 17. Available at http://spie.org/newsroom/2503-improved-pedestrian-tracking-for-urban-planning.

Silver, Phil. 2015. "Urbanomics Mobility Project." *Urban Insights*, October 9. Available at www.urban-insights.com/Blog/Articles/ID/55/Urbanomics-Mobility-Project.

Smith, Kendra. and Kevin Desouza. 2015. "How Data Privatization Will Change Planning Practice." *Planetizen*, July 20. Available at www.planetizen.com/node/79680/how-data-privatization-will-change-planning-practice.

Smith, Kendra L., Isabel Ramos, and Kevin C. Desouza. 2015. "Economic Resilience and Crowdsourcing Platforms." *Journal of Information Systems and Technology Management* 12 (3): 595–626.

Sopkin, Kristin L., Hilary F. Stockdon, Kara S. Doran, Nathaniel G. Plant, Karen LM Morgan, Kristy K. Guy, and Kathryn EL Smith. 2014. *Hurricane Sandy: Observations and Analysis of Coastal Change*. Washington, DC: US Geological Survey. Available at http://pubs.usgs.gov/of/2014/1088/pdf/ofr2014-1088.pdf.

Spector, Julian. 2015. "How Portable Air Sensors Are Changing Pollution Detection." *CityLab*, August 13. Available at http://www.citylab.com/weather/2015/08/how-portable-air-sensors-are-changing-pollution-detection/401147/.

Sterman, John D. 2000. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. New York: Irwin/McGraw-Hill.

Symantec. 2015. *Internet Security Threat Report 20*. Available at www.symantec.com/security_response.

Tabbitt, Sue. 2014. "Big Data Analytics Keeps Dublin Moving." *Telegraph*, February 17. Available at www.telegraph.co.uk/sponsored/sport/rugby-trytracker/10630406/ibm-big-data-analytics-dublin.html.

Talbot, David. 2013. "African Bus Routes Redrawn Using Cell-Phone Data." *MIT Technology Review*, April 30. Available at www.technologyreview.com/news/514211/african-bus-routes-redrawn-using-cell-phone-data.

Thakuriah, Piyushimita, Nebiyou Tilahun, and Moira Zellner. 2015. "Big Data and Urban Informatics: Innovations and Challenges to Urban Planning and Knowledge Discovery." In *Proceedings of the Workshop on Big Data and Urban Informatics*, 4–32, Chicago, IL, August 11–12.

Thaler, Richard H., and Cass R. Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. 2008. New Haven, CT: Yale University Press.

Tollefson, Jeff. 2015. "Can Randomized Trials Eliminate Global Poverty?" *Nature* 524 (7564): 150–153.

Transport for London. 2016. "Ultra Low Emission Zone." Available at https://tfl.gov.uk/modes/driving/ultra-low-emission-zone.

Treasury Inspector General for Tax Administration. 2015. *The Return Review Program Enhances the Identification of Fraud; However, System Security Needs Improvement*. Available at www.treasury.gov/tigta/auditreports/2015reports/201520060fr.pdf.

Twitter. 2016. "It's What's Happening." Available at https://about.twitter.com/company.

University of Utah. 2015. "Programming and Prejudice." *UNews*, August 14. Available at http://unews.utah.edu/programming-and-prejudice.

UPS. 2016. "ORION Backgrounder." Available at www.pressroom.ups.com/pressroom/ContentDetailsViewer.page?ConceptType=Factsheets&id=1426321616277-282.

Urban, Glen L., and Eric Von Hippel. 1988. "Lead User Analyses for the Development of New Industrial Products." *Management Science* 34 (5): 569–582.

Urmson, Chris. 2015. "Chris Urmson, Google, Inc. November 12, 2015 Response Letter." Available at http://isearch.nhtsa.gov/files/Google%20--%20compiled%20response%20to%2012%20Nov%20%2015%20interp%20request%20--%204%20Feb%2016%20final.htm.

US Department of Justice. 2015. *Investigation of the Ferguson Police Department*. Available at www.justice.gov/sites/default/files/opa/press-releases/attachments/2015/03/04/ferguson_police_department_report.pdf.

Vallianatos, Mark. 2015. "How LA Used Big Data to Build a Smart City in the 1970s." *Gizmodo*, June 22. Available at http://gizmodo.com/uncovering-the-early-history-of-big-data-in-1974-los-an-1712551686.

Vigen, Tyler. 2015. *Spurious Correlations*. New York: Hachette Books.

Von Hippel, Eric. 1986. "Lead Users: A Source of Novel Product Concepts." *Management Science* 32 (7): 791–805.

Ward, Mark. 2014. "Smart Meters Can Be Hacked to Cut Power Bills." *BBC News*, October 16. Available at www.bbc.com/news/technology-29643276.

Warwick Business School. 2015. "Twitter and Mobile Phoes Used to Calculate Crowd Sizes." WBS News, May 27. Available at www.wbs.ac.uk/news/twitter-and-mobile-phones-used-to-calculate-crowd-sizes.

West, Geoffrey. 2013. "Big Data Needs a Big Theory to Go with It." *Scientific American*, May 1. Available at www.scientificamerican.com/article/big-data-needs-big-theory.

Wilensky, Uri, and William Rand. 2015. *An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo*. Cambridge, MA: MIT Press.

Wogan, J. B. 2015. "How Mobile, Alabama, Used Instagram to Address Blight." *Governing*, November 20. Available at www.governing.com/topics/transportation-infrastructure/gov-mobile-alabama-blight-instagram.html.

Zavattaro, Staci M., P. Edward French, and Somya D. Mohanty. 2015. "A Sentiment Analysis of US Local Government Tweets: The Connection between Tone and Citizen Involvement." *Government Information Quarterly* 32 (3): 333–341.

Zheng, Hong, Young-Jun Son, Yi-Chang Chiu, Larry Head, Yiheng Feng, Hui Xi, Sojung Kim, and Mark Hickman. 2013. *A Primer for Agent-Based Simulation and Modeling in Transportation Applications*. Washington, DC: US Department of Transportation. Available at www.fhwa.dot.gov/advancedresearch/pubs/13054/13054.pdf.

ACKNOWLEDGMENTS

The authors would like to thank the individuals who contributed to the development of this report. **Lindsey Collins**, Arizona State University, assisted with research on all chapters. Her time, effort, and enthusiasm for this project were invaluable. **Rashmi Krishnamurthy**, Arizona State University, and **Srinivasa Srivatsav Kandala**, Arizona State University, also assisted with the development of segments of this report. We appreciate the time and attention of the reviewers and contributors who provided meaningful feedback and content to enrich this effort: **Nader Afzalan**, American Planning Association and University of Redlands; **Tony Grubisec**, Arizona State University; **Spiro Maroulis**, Arizona State University; and **Shannon McElvaney**, Esri.

American Planning Association
205 N. Michigan Ave., Suite 1200
Chicago, IL 60601-5927

planning.org