URBAN ANALYTICS

SPATIAL ANALYTICS AND GIS SERIES

In the digital age, social and environmental scientists have more spatial data at their fingertips than ever before. But how do we capture this data, analyse and display it, and most importantly, how can it be used to study the world?

Spatial Analytics and GIS is a series of books that deal with potentially tricky technical content in a way that is accessible, usable and useful. Early titles include **Urban Analytics** by Alex Singleton, Seth Spielman and David Folch, and **GIS and Agent-based Modelling** by Andrew Crooks, Nick Malleson and Alison Heppenstall.

Series Editor: Richard Harris

About the Series Editor

Richard Harris is Professor of Quantitative Social Geography at the School of Geographical Sciences, University of Bristol. He is the lead author on three text-books about quantitative methods in geography and related disciplines, including *Quantitative Geography: The Basics* (SAGE, 2016).

Richard's interests are in the geographies of education and the education of geog-raphers. He is currently Director of the University of Bristol Q-Step Centre, part of a multimillion pound UK initiative to raise quantitative skills training among social science students, and is working with the Royal Geographical Society (with IBG) to support data skills in schools.

Books in this Series:

An Introduction to Big Data and Spatial Data Analytics in R, Lex Comber & Chris Brunsdon

An Introduction to R for Spatial Analysis and Mapping, 2nd Edition, Chris Brunsdon & Lex Comber

Geocomputation, Chris Brunsdon & Alex Singleton

GIS and Agent-based Modelling, Andrew Crooks, Nicolas Malleson & Alison Heppenstall

Modelling Environmental Change, Colin Robertson

Published in Association with this Series:

Quantitative Geography, Richard Harris

URBAN ANALYTICS

ALEX D. SINGLETON
SETH E. SPIELMAN
DAVID C. FOLCH

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC | Melbourne

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC | Melbourne

SAGE Publications Ltd
1 Oliver's Yard
55 City Road
London EC1Y 1SP

SAGE Publications Inc.
2455 Teller Road
Thousand Oaks, California 91320

SAGE Publications India Pvt Ltd
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road
New Delhi 110 044

SAGE Publications Asia-Pacific Pte Ltd
3 Church Street
#10-04 Samsung Hub
Singapore 049483

Editor: Robert Rojek
Editorial assistant: Catriona McMullen
Production editor: Katherine Haw
Copyeditor: Neville Hankins
Proofreader: Lynda Watson
Indexer: Cathryn Pritchard
Marketing manager: Susheel Gokarakonda
Cover design: Wendy Scott
Typeset by: C&M Digitals (P) Ltd, Chennai, India
Printed in the UK

First published 2018

Library of Congress Control Number: 2017941647

British Library Cataloguing in Publication data

A catalogue record for this book is available from the British Library

ISBN 978-1-4739-5862-3
ISBN 978-1-4739-5863-0 (pbk)

1: QUESTIONING THE CITY THROUGH URBAN ANALYTICS

Learning Objectives

By the end of this chapter students will understand the following:

- The majority of future population growth will be concentrated in urban areas.
- The planning and management of population change creates a range of challenges for cities.
- New forms of technology are increasingly embedded into city systems and are providing a wealth of new data.
- Urban analytics represents a methodological toolkit for studying and managing data-rich cities.

Human Settlements and Urbanization

Looking down onto the surface of the earth from space reveals a patchwork of urbanization. By 2015 the United Nations (UN) estimated that global population reached 7.3 billion people, and projected that this will have grown to 9.7 billion by 2050 (United Nations 2015). Not only has the world's population grown, but the distribution of people has changed over time, and at some point in 2009 we reached a significant tipping point where more than half of the global population lived within urban as opposed to rural areas (United Nations 2009). This transition to urbanized living is driven by complex economic, technological, cultural, and geopolitical forces. Global urbanization has fundamentally changed how we (as a species) interact with one another and our natural environment, creating a range of challenges for the planning and management of urban areas.

By night, the scale and extent of urban areas can be rendered visible through satellite imagery that records the use of electric lighting (see Figure 1.1). However, the social, economic, and environmental impacts of these brightly lit places extend far beyond these borders. From space, those dark areas between city lights seem empty and undifferentiated. One can imagine these dark spaces to be a patchwork of agricultural and natural areas that are disconnected from the networks of lights. However, much of the globe, even these dark places, are linked together through a network of ecological and economic exchanges that fuel and support urbanization. Much of this network is material – a vast communications infrastructure that, while concentrated in cities, spans the entire globe.

What has fueled this growth? There is not one story – the forces that drive urbanization in the more developed parts of the world are different from the forces driving urbanization in the global South. Generically, residents of cities are attracted by the advantages of proximity to sources of employment, infrastructure, and cultural assets, or perhaps as a result of improved provision of healthcare or sanitation. The net result of this growth of cities is that a large share of the world's population is connected to the economic and technological infrastructure that emanates from them.

The phenomenon of urbanization is not new, nor are those challenges emerging from these processes. However, a critical difference between urbanization of the twenty-first century and the waves of urbanization that have occurred in the past is that there are entirely new ways to understand these processes. The same information technologies that connect cities in a global network can also be used within them to manage

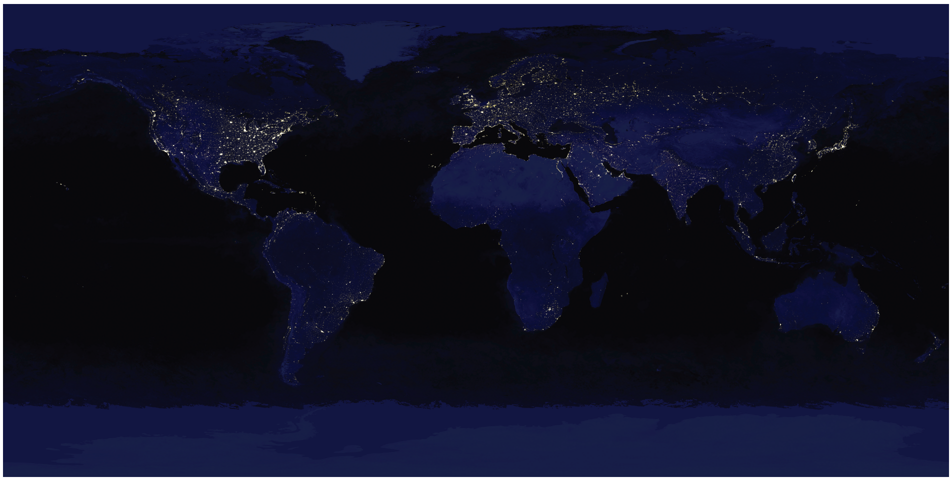

Figure 1.1 At night, urbanization is rendered visible on the earth's surface through a satellite originally designed to detect cloud coverage

Source: Data courtesy Marc Imhoff of NASA GSFC and Christopher Elvidge of NOAA NGDC. Image by Craig Mayhew and Robert Simmon, NASA GSFC the provision of services, and to mitigate the environmental impact of their metabolism within and beyond their borders. New ways of knowing and managing cities are occurring because of advances in those instruments that can monitor activities within or attributes of urban environments (see Chapter 2). Enhancements to communications infrastructure are enabling the data generated by such devices to be utilized by services in real time, and for devices to communicate with one another, potentially making automated decisions based upon derived information.

The macro shape, structure, and function of urban areas emerge through an incredibly complex set of human interactions; cities are dynamic systems that evolve from the bottom up and over time (Batty 2013). Currently the most urbanized regions of the world include North America, Europe, Latin America, and the Caribbean (see Table 1.1); however, by 2050, it is estimated that 37 percent of new projected urban population growth will be attributable to just three countries in Africa and Asia, namely, China, India, and Nigeria (United Nations 2015).

Table 1.1 Regions of the world vary significantly by their levels of urbanization (United Nations 2015)

Region	Urban	Rural	Total	Percentage urban
Africa	455,345	682,885	1,138,229	40.0
Asia	2,064,211	2,278,044	4,342,255	47.5
Europe	545,382	197,431	742,813	73.4
Latin America and the Caribbean	495,857	127,565	623,422	79.5
North America	291,860	66,376	358,236	81.5
Oceania	27,473	11,356	38,829	70.8
World	3,880,128	3,363,656	7,243,784	53.6

Note: Population is in thousands.

Source: United Nations, World Urbanization Prospects, 2015

However, the global narrative of increasing urbanization does not hold everywhere. While some cities expand rapidly, others are shrinking in response to macroeconomic trends, like the decline of manufacturing in many developed countries, environmental disasters (Box 1.1), and/or political instability. Urban areas require careful management, planning and investment if the negative societal impacts that can be associated with rapid or long-term change are to be avoided or mitigated. For cities with declining population, sustaining infrastructure (e.g., roads or schools) that was designed to service much larger populations can strain municipal finances and make it difficult to provide basic services like education and safety. Conversely, in a place with rapid population growth, how might public transit systems be reconfigured to meet this extra demand given land use constraints, such as increasing the density of people's homes located at the site of a potential train stop, that might prevent the expansion of the transit network?

Governing cities is a complex and political process. While the new forms of data about cities do not ease the political burden, they can feed into decision-making processes and help to manage existing infrastructure more efficiently. Decisions about cities are complex, involving difficult trade-offs. However, there are a wealth of often disparate empirical data, a suite of methods or tools for translating this into information, and a mechanism for the communication of findings to stakeholders (Longley et al. 2015b). This book provides a background to this process.

BOX 1.1

Change in Urban Areas – Hurricane Katrina and Information Technology

Hurricane Katrina struck the Gulf Coast of the United States in August 2005, and had devastating impacts on the population, infrastructure, and economy of the region. The city of New Orleans, Louisiana was acutely affected after a number of levees failed, leaving 80 percent of the city flooded under 15 to 20 feet (4.5 to 6 m) of water (Figure 1.2).

Figure 1.2 Searching for survivors in New Orleans

Source: US Coast Guard photo by Petty Officer 2nd Class NyxoLyno Cangemi (Wikipedia)

(Continued)

The total costs of Hurricane Katrina were catastrophic, with the National Oceanic and Atmospheric Administration (NOAA) estimating that there were 1,353 direct fatalities, 275,000 homes damaged or destroyed, along with financial costs in excess of $100 billion (Johnson 2006). With such significant impact to housing infrastructure and property, this resulted in a huge number of people being displaced into new locations. Figure 1.3 is not a direct proxy of these movements, but illustrates change at a state level after this event.

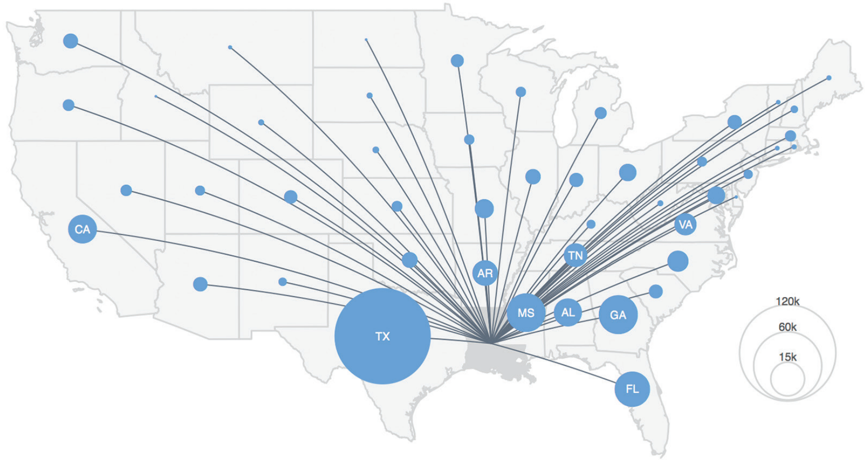

Figure 1.3 States where migrants housed within Louisiana during 2005 had moved to as of 2006

Source: Authors' own; data – US Census Bureau, American Community Survey 2006; State to State Migration Flows 2004–2015

The impact of Hurricane Katrina on New Orleans illustrates the dynamic nature of urban areas, and indeed, over a decade on from the event, population levels have not recovered fully – although, even prior to the hurricane, the population had also been in decline from a peak in the 1970s (Figure 1.4).

In the aftermath of Hurricane Katrina, the then nascent technologies from Google, including its Maps and Earth platforms, were used by both official bodies and the public to collect or disseminate spatial data related to the event. This included the loading of high-resolution aerial imagery into Google Earth

(Continued)

(Continued)

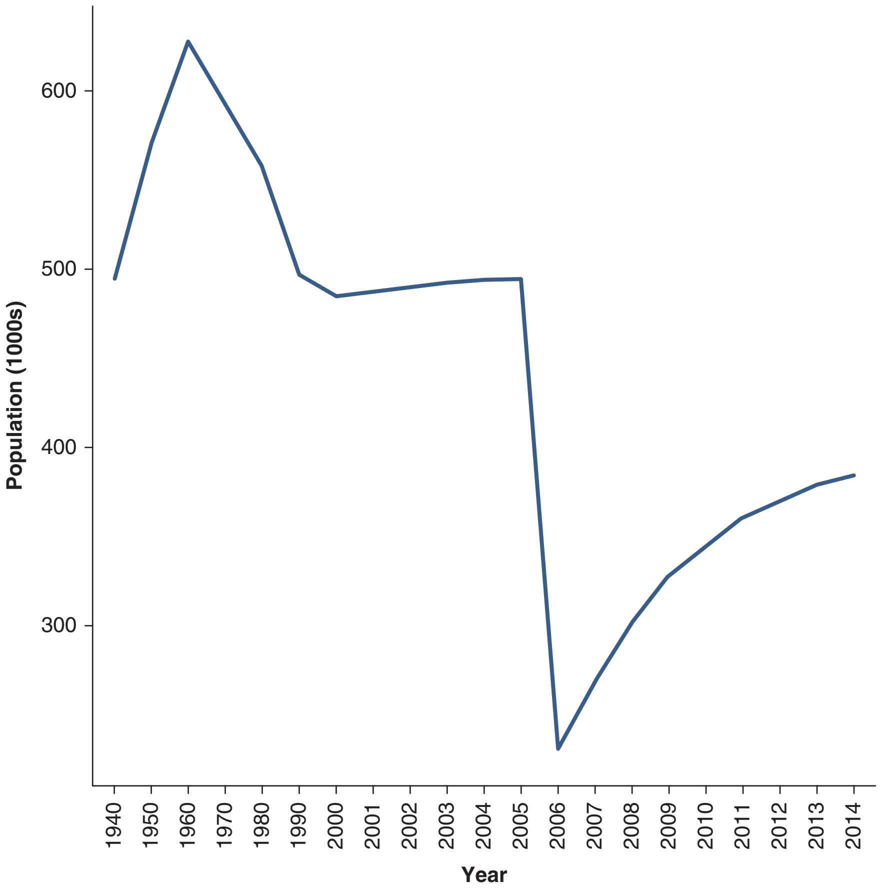

Figure 1.4 Population change in New Orleans between 1950 and 2014

Source: Authors' own; data – US Census Bureau, Population of the Largest 100 Urban Places 1940–2016

by NOAA, and additionally an early grassroots example of a map "mashup" where survivors posted details of their locations or other useful information (Figure 1.5). The relationship between technology platforms, urban areas, and their inhabitants is an enduring feature of contemporary urban living, and a key facet of urban analytics relates to the use of these new spatial data infrastructures for decision making.

(Continued)

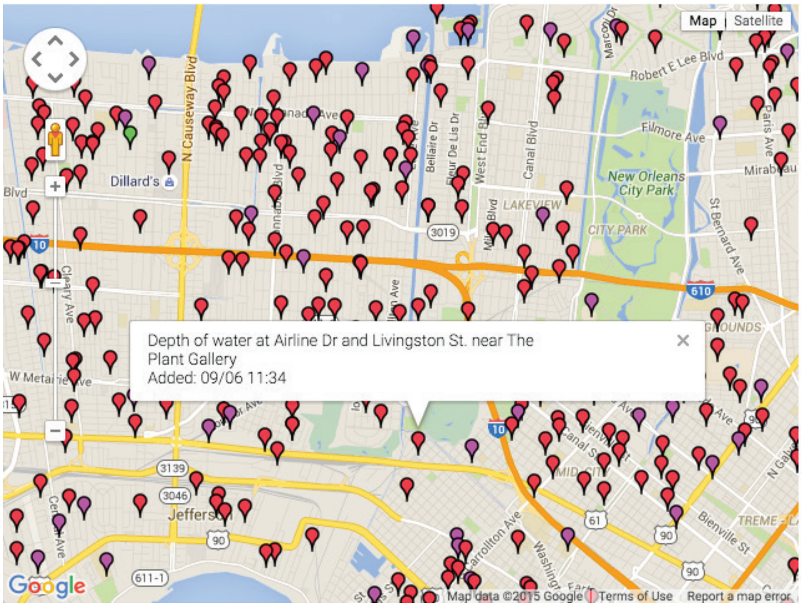

Figure 1.5 An early map "mashup" created in the aftermath of Hurricane Katrina

Source: Website was developed by Jonathan Mendez and Greg Stoll and an archive of the site is available online (http://www.gregstoll.com/scipionus)

Urban Data Systems

New York City (NYC) is situated on the north-eastern seaboard of the United States of America, and is the country's most populated city, with the 2010 Census recording over 8 million residents. As with many large metropolitan areas across the globe, technology and data are increasingly embedded into the fabric of the urban area making up NYC.

Figure 1.6 shows a photograph of a fairly typical street scene occurring within Times Square, which is located within the Mid-Town district on the island of Manhattan. People can be seen intermingling with traffic, and theaters, cinemas, and shops line the pedestrianized area, with residential and office towers dominating the skyline. However, beyond the obvious visible features, various technologies infiltrate multiple aspects of urban living and mobility.

Way beyond the visible range of the optics used on the camera that took this photograph, and at approximately 20,000 km above the surface of the earth, a constellation of Global Positioning System (GPS) satellites orbit. Back in Times Square, and among the crowd of pedestrians, a family of tourists has become disorientated; fortunately, however, they have a GPS-enabled cell phone which combines information returned from the satellites with data from a technology company in California to calculate their location. This is then plotted on a map, and routing functionality assists them to navigate to their intended destination. The maps and route calculation are provided over a wireless Internet connection which is also being used by hundreds of other people within this area. Many of these people have security settings on their cell phones that enable the sharing of locational information collected by a range of technology companies. One provider of free public maps collects this data, pooling the locations, times, and speeds of users within this area, and uses these attributes in a model that predicts traffic congestion. Based on this real-time information, the taxi cab just pulling out of the photograph has rerouted to avoid additional traffic that will cause delay to its current fare.

On top of a number of lampposts are security cameras streaming live video over the Internet. A shop owner, watching the video, identifies an instance of petty theft and the NYC Police are dispatched. The perpetrator is arrested with the location, time, and offense recorded. These attributes are later integrated into extracts placed on the NYC Open Data store. Researchers at a university then use these data to develop a statistical model that estimates where and when crime is most likely to occur, which later feeds into a revised policing strategy report.

Many of the people in the photograph are holding shopping bags, having made purchases from the stores running along the side of Times Square, and typically have paid for these goods using credit cards. After the transactions are processed, records of these purchases are added to the flow from millions of other transactions by consumers from across the United States, with this data being stored by the credit card company. The updated records feed dynamically into the consumer segmentation models which are then provided as a service to third-party marketing agencies.

Collectively all of these small interactions can have a profound impact on the state of a city at any given moment in time. Consider that taxi, using a routing application to avoid traffic. If 10 percent of drivers are using such an application in an area, a significant portion of traffic will be redirected from one part of the city to another. However, such routing applications optimize a single person's route, but when lots of people are using them they have an unobservable impact on the fabric of a city. The wealth of data generated within or used by urban areas is not restricted to these brief examples; however, they aim to be illustrative of the ways in which data flows through urban systems and may be repurposed. These small technology-mediated interactions have a growing role in the shape of urban life. Increasingly cities are directly supporting digital infrastructure as they would traditional "hard" infrastructure like roads. In NYC, for example, public pay telephones are being replaced with LinkNYC (www.link.nyc), which will provide a digital communications infrastructure, including support for navigation, Wi-Fi access, and device charging (Figure 1.7).

Figure 1.6 Times Square, NYC: technology is integral to how cities can be managed and studied

Source: Photograph by Aurelien Guichard CC BY-SA (Flickr)

Figure 1.7 LinkNYC Station – reimagination of the telephone booth

Source: Seth Spielman

BOX 1.2

Urban Big Data

When talking about cities it is difficult to escape the term Big Data, which has become commonplace in the description of a particular aspect of the evolving data economy and its links to servicing infrastructure. There are numerous definitions of Big Data (Kitchin and McArdle 2016); however, many have typically made reference to three main "Vs" which include: volume, velocity, and variety, although these are increasingly supplemented by other characteristics (see Table 1.2). Not all data related to urban areas are Big Data, and the majority of data featured within this book are typically large in size, but do not fit a formal definition of Big Data.

Table 1.2 The characteristics of Big Data

Characteristic	Description
Volume	Huge in *volume*, consisting of terabytes or petabytes of data
Velocity	High in *velocity*, being created in or near real time
Variety	Diverse in *variety* in type, being structured and unstructured in nature, and often temporally and spatially referenced
Exhaustive	*Exhaustive* in scope, striving to capture entire populations or systems (n = all), or at least much larger sample sizes than would be employed in traditional, small data studies
Resolution	Fine-grained in *resolution*, aiming to be as detailed as possible, and uniquely indexical in identification
Relational	Containing common fields that enable *relational* joins to different datasets
Flexible	Structured in a *flexible* way that enables new fields to be easily added
Scalable	Fully *scalable* and can be expanded rapidly

Source: Adapted from Kitchin (2014: 68)

Multiple factors have accelerated the availability of data within an urban setting, and this creates a demand for professionals trained in analytics and urban planning to effectively use such new resources. Until recently, access to urban data was often difficult; however, a trend toward more open licensing of data by cities enables use of municipal data without financial cost. The emergence of central repositories that consolidate data for a city facilitate discovery and download of these assets. An example of such a project is outlined in Profile 1.1.

Profile 1.1

Charlie Catlett

Figure 1.8 Charlie Catlett, Director, Urban Center for Computation and Data, University of Chicago

Source: Supplied by Charlie Catlett, image copyright © Eileen Moloney

Describe how you got interested in urban data science and your labs research

Over the longer term my research interests have concerned various aspects of the Internet, distributed and high-performance computing; however, about five years ago I became particularly interested in the rapid growth of cities, and in particular those huge investments happening and will happen over the next few decades in new urban infrastructure. So my interest started by asking if we have the right design tools and understanding of cities to develop infrastructure over the next 50 years that will be sustainable, and will correct some of our past mistakes. Through conversations with the administration of numerous cities over the course of a few years, it became clear to me that

(Continued)

(Continued)

having a measurement system like the Array of Things (arrayofthings.github.io) would be useful to the science community to enable them to work with residents and city organizations or governments to address neighborhood challenges.

In general, within our center we have three areas of research related to cities' urban data science. The first concerns the development of tailored applications and tools that aim to make Open Data more accessible to the science community and to the general public. The second area is to develop new capabilities that use open and other data to better understand neighborhoods and cities, particularly in terms of their economic resilience, public safety, education, and health. This was motivated through early discussion with the City of Chicago Mayors Office who had interest in examining neighborhood trends over time; and specifically to identify where interventions may be necessary before a neighborhood reaches a crisis point. Finally, we have interests in embedding systems within cities to support research concerning the development of key capabilities in both measurement and actuation.

How can intelligent infrastructure help cities?

Intelligent infrastructure can really help by reducing the time frame within which you can make a decision, or examine how the result of that decision manifest. For many decisions within cities, such as the design of a new highway, park, or installation of a new rapid bus transit, their design and implementation may unfold over several years, and their impact on a city potentially over several decades. For such applications, historical data and simulation of potential outcomes are important; however, there are other areas that consider decisions for much higher temporal resolutions, for example, traffic safety. Within these examples, embedded systems become important, and especially so when coupled with capability to make decisions in real time.

What is the Array of Things?

This started off as a connected sensor network project focused on the urban environment; however, it has evolved into a project about embedded capabilities to support research over multiple areas. Rather than try to develop a set of particular capabilities like controlling street lights or an application that tells you about air quality, we instead have focused on a general platform that allows for innovation and the rapid deployment of new hardware and software within the city. We also do this while making

(Continued)

data available in forms and through mechanisms that will encourage people to innovate in terms of analytics and application development; but, and very importantly, in a way that preserves the privacy of individuals.

How can you reassure citizens about the collection and use of urban sensor data?

One of the things we have done was to develop a set of clear messages about the Array of Things project and take these out into public meetings where we have a chance to interact with people who are interested and can spend time answering questions. The bottom line for most people is that they want to know what you are doing with the data and how you are ensuring privacy. However, what we have learnt within Chicago is that clarity of communication and transparency, although really important, aren't entirely sufficient. As such, the third area that we have developed as part of our policies is structured and transparent accountability, that we ensure through an independent external privacy committee whose role is explicitly embedded within our governance policies. Where we wish to make changes to our underlying technology (e.g., a new sensor), we are required to make a report to the committee about the scientific justification for a change, the privacy implications that will come about because of them, and what actions we will take to continue to preserve privacy. They may have feedback or recommendations which are given to our governance committee, with these transactions made public.

These developments have required engagement from data owners such as government departments, who invest time in preparing data in formats that make them suitable for wider use. Given the cost of municipal data collection and curation, not all cities are able to offer such resources. In addition to city-to-city variation within countries, practices of municipal data dissemination are variable between countries. For example, within the United States there is a history of data collected by public agencies being placed into the public domain; for example, the decennial census of the population. However, this is not the case everywhere; within the UK, for example, prior to 2001, census data were only made available through a restricted set of commercial resellers, despite the collection of the data being publicly funded.

The arguments for making data open have both an ethical and a pragmatic dimension. Ethically, if the public have funded the collection of a particular dataset (e.g., through taxes), then it could be argued that this should also be made available publicly, with the caveat of appropriate disclosure controls or security constraints. Secondly, the release of Open Data has the potential to generate large economic or societal benefit. In 2013, the global consulting firm McKinsey estimated that worldwide there were approximately $3 trillion worth of value contained within Open Data (Manyika et al. 2013). The availability of Open Data has created new business opportunities, with organizations reselling value-added versions of the data, or integrating the data into new products or services. Transparency in the collection and dissemination of urban data is a good thing. When cities expose to citizens (and visitors) the kinds of information they collect, people are more informed about how their activities are monitored. However, simply making data available does not lead to its effective use. For the value identified by McKinsey to be materialized, creative and innovative use of data are necessary. Understanding how to effectively use urban data is a core goal of this book.

As the volumes of Open Data have grown, a number of Open Data platforms have emerged that provide search and discovery of resources. The two most prevalent systems implemented include CKAN (ckan.org) and Socrata (www.socrata.com), and platforms based on these systems have been developed for a variety of national extents (see Table 1.3), or within the context of specific regions or cities (see Figure 1.9).

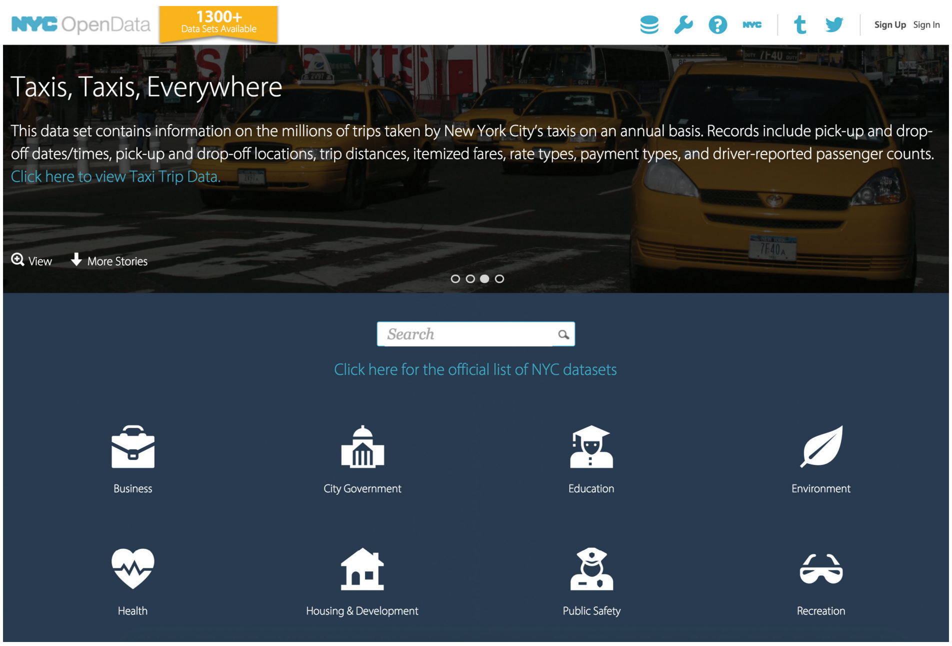

Figure 1.9 NYC OpenData. Many cities across the world now have datastores that provide access to Open Data related to multiple attributes about cities

Source: https://nycopendata.socrata.com/. Reprinted with permission

Table 1.3 Some examples of national Open Data portals

Country	Web address	Platform
UK	data.gov.uk	CKAN
Japan	www.data.go.jp	CKAN
Kenya	www.opendata.go.ke	Socrata
United States	data.gov	CKAN
India	data.gov.in	Open Government Platform

Practicing Urban Analytics

The widespread availability of urban data is a new phenomenon, and creates potential for a new field of inquiry called "urban analytics." Urban analytics is the practice of using new forms of data in combination with computational approaches to gain insight into urban processes. Increasing data availability allows us to ask

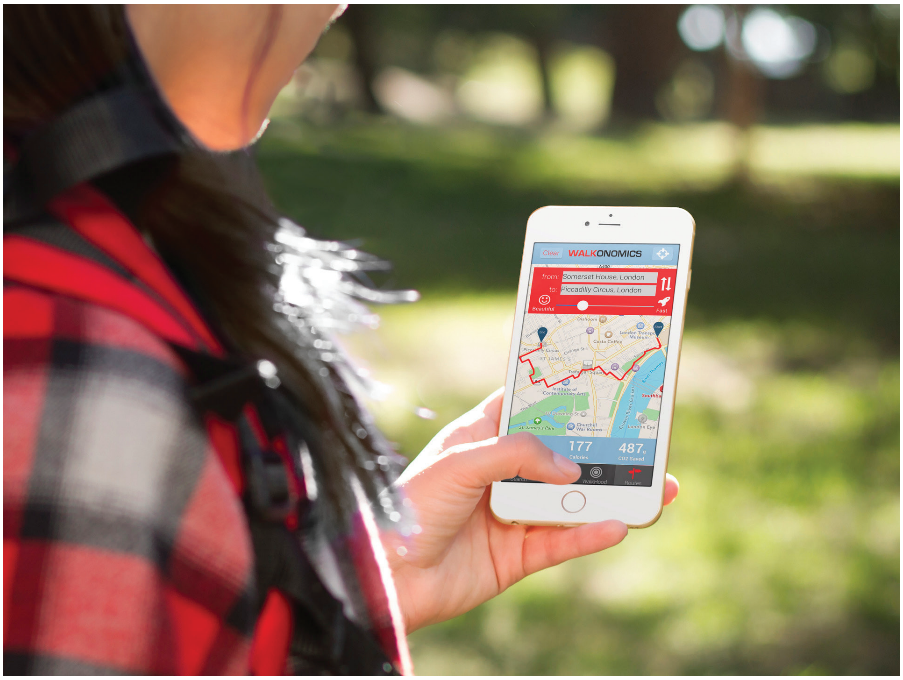

Figure 1.10 Walkonomics use a variety of data about cities to grade streets by a series of measures that impact the pedestrian experience. These can be integrated into their mobile application that enables routing

Source: Image available at data.gov.uk. Walkonomics app, www.walkonomics.com/

new and often complex questions about cities, their economy, how they relate to the local and global environment, and much more. For example, something as mundane as adjusting the timing of traffic signals in response to fluxes in vehicle volume derived though sensor networks could dramatically reduce carbon emissions in a city. Providing automated alerts to mobile devices if air quality were to fall bellow an acceptable threshold could improve public health and awareness of pollution problems. In a more traditional sense, these new data can also be integrated to improve strategic planning systems, by providing a richer evidence base upon which to make more optimal decisions given constraints. Finally, data can empower citizens to make more informed decisions about its use, mobility, and governance of urban areas (Figure 1.10).

Data in itself does not offer insight, and methods are required to both generate and communicate findings. Urban analytics provides the tools, technologies, and processes for the pursuit of this new data-intensive science of cities. However, even when we ask the right questions with an appropriate method, we can potentially get incorrect answers. With most new data we may trade the breadth and geographic scale of the attributes being measured for incompleteness or uncertainty. The overarching objective of this book is to introduce the main techniques of urban analytics, their underlying science, and provide consideration of their appropriate use.

Questions

1. Where are the largest cities in the world and how can these be measured? Think about the different ways in which we might measure "large" and explore what data are available to derive this information.

2. Explore changes in city population over time for a country of your choice. Describe the patterns with reference to short-term (e.g., environmental disaster, political instability, etc.) and long-term stimulus (e.g., economic decline etc.).

3. Think about a city near you or the one in which you live. Write a 2,500-word essay on the different ways in which data are embedded into the urban environment, making reference to how ethical considerations in the use of the data can be balanced against social, environmental, or economic benefits.

4. Compare and contrast two city datastores, making reference to the breadth, depth, and scale of data that they hold. How might these datastores be improved?

Supplementary Reading

Batty, M. (2013). *The new science of cities*. Cambridge, MA: MIT Press.
This book provides a window onto a lifetime of experience in urban and regional modeling, and makes the case for a new science of cities. It is an advanced text, but very readable, even by those who are new to this area.

LeGates, R. T., & Stout, F. (2015). *The city reader*. Abingdon: Routledge.
One of the best anthologies of academic writing on cities, planning, and urban studies more generally, mixing a range of classic and contemporary writings from key thinkers. The contents are mainly theoretical with many of the writings also providing rich historical accounts.

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2015). *Geographic information science and systems* (4th edition). Hoboken, NJ: Wiley.
This book is a classic reference text for GISc and provides excellent background materials to this chapter and the rest of the book more generally.