Chapter 21

Data Integration for Urban Health

Yuan Lai and David J. Stone

Abstract Population wellness and quality of life are the results of complex urban biophysical and socioeconomic dynamics. This article explores this novel domain and introduces a broader context of urban data landscape and health indications. It also provides a specific project narrative as an exemplar—NYC PollenScape—describing its background, data, methods, and findings. A more general discussion of sociotechnical challenges in integrated data analytics for urban health follows. The paper concludes with a summary of key takeaways, current limitations, and future work.

Keywords Urban planning · Smart cities · Digital health · Population health · Machine learning · System engineering · Socio-technical model

Learning Objectives

By the end of this chapter, you will be able to:

1. Present the current urban data landscape including sources, typology, and limitations.
2. Describe health use cases of urban data.
3. Understand the socio-technical considerations and challenges around data integration for urban health.

21.1 Introduction

Since the 1940s, scientific research in cities has evolved with the inclusion of ideas from regional science (Zipf 1949), cybernetics (Wiener 1948), systems engineering

---

Y. Lai (✉)

Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

e-mail: yuanlai@mit.edu

D. J. Stone

Departments of Anesthesiology and Neurosurgery, University of Virginia School of Medicine, Charlottesville, VA, USA

© The Author(s) 2020

L. A. Celi et al. (eds.), *Leveraging Data Science for Global Health*,
https://doi.org/10.1007/978-3-030-47994-7_21

351

(Von Bertalanffy 1969), and system dynamics (Forrester 1970). These studies have conceptualized cities as complex ‘systems-of-systems’, albeit with limitations due to a lack of real-world data and the hardware and software elements required for implementations. Over the past two decades, rapid technological development has brought smart devices, the quantified-self movement (Wolf 2007), the smart cities movement (Wiig 2015), big data (James et al. 2011), and the open data movement (Barbosa et al. 2014) into the urban studies realm. As a result, we are becoming “human on the net” (Bradley 2007)—well-connected and integrated with urban sociotechnical systems. Meanwhile, rapid urbanization, population growth, and climate change are intensifying at an unprecedented speed. Citizens, especially vulnerable populations, are facing threats to achieving ecological sustainability, economic prosperity, social justice, and overall quality of life. Emerging disciplines, such as urban computing, urban informatics, and civic analytics, have begun to incorporate data science and urban domain knowledge in order to generate a more scientific and consistent understanding of cities. Such efforts have generated two main branches of research output: technological deployment within an urban environment (“smart cities”) and the new insights derived from real-world data that constitute “urban intelligence”. Regarding the first branch, urban infrastructure is increasingly equipped with real-time sensing nodes, cloud computing, automation systems, and optimized networks. And regarding the second, ubiquitous urban data increasingly enable data-supported policy formulation, computational social science, preventive interventions, and data-driven operations.

This article maintains that data integration is a critical component for developing next-generation intelligence for urban health, and that population wellness and quality of life are the results of complex urban biophysical and socioeconomic dynamics. After introducing a broader context of urban data landscape and health indications, it then provides a specific project narrative as an exemplar—NYC Pollen-Scape—describing its background, data, methods, and findings. A more general discussion of sociotechnical challenges in integrated data analytics for urban health follows. The paper concludes with a summary of key takeaways, current limitations, and future work.

21.2 The Broader Context

21.2.1 Urban Data Landscape

Urban data are large in volume and variety, and derive from heterogeneous sources that require different analytical strategies. Conventionally, government administrative data at the federal, state, and city level provide baseline information for population and neighborhood studies. The U.S. Census Bureau conducts nation-wide

decennial population censuses, and provides data to support large-scale decision-making and policy analysis involving housing, transportation, healthcare, and education (U.S. Census Bureau 2010). The American Community Survey is an annual sampled survey reporting neighborhood population characteristics regarding demographics, occupation, income, and education, in order to inform public funding, capital planning, and infrastructure investment (U.S. Census Bureau 2018). New York City Community Health Profiles report neighborhood-level health conditions (life expectancy, healthcare, behavioral and demographic characteristics) from fifty-nine community districts in New York City (NYC Department of Health and Mental Hygiene 2018). The population census and community survey data provide long-term, comprehensive information to support cross-sectional longitudinal analyses. Universities, research institutes, and non-profit organizations represent additional sources providing database archives, data repository portals, and online platforms related to global health. These platforms enable researchers and health practitioners to explore comprehensive information collection. However, since these platforms collect data from various sources and share access as a third party, the timeliness and coverage of available resources are highly dependent on platform maintenance.

With increasing digital applications and IoT (Internet of Things) products, Application Program Interfaces (API) have become a relatively new data source. *AIDSinfo* is an API created by the U.S. Department of Health and Human Services for accessing AIDS-related drug databases (U.S. Department of Health and Human Services 2019). *AirNow* is an API providing real-time air quality data in the U.S., Canada, and Mexico (US EPA 2019). Fitbit is a wearable device for tracking personal physical activities (walking and sleeping) and health (heart rate), providing an API for accessing the data (FitBit 2019). Twitter, one of the most popular online news and social media platforms, provides a developer platform with multiple APIs that enable the query of historical data (most recent seven days) by keywords, the creation of a campaign, or the generation of social media engagement metrics (Twitter 2019). All of these APIs provide data resources for prototyping new applications or integrated products, which also require deep technical skills to access, read, and process data.

21.2.2 Urban Health Indications

Increasing data sources and multidisciplinary research methods have enabled new analytic approaches to public health research and practice. These include historical clinical data, hospitalization records, and surveys that report certain aspects of health outcome (e.g., mortality rate, life expectancy, hospitalization rate, asthma, diabetes). The United States Small-Area Life Expectancy Estimates Project (USALEEP) is the most granular (specifically at the census tract level) and comprehensive (all neighborhoods in the U.S.) study to date (NAPHSIS 2018). This program reveals that in addition to healthcare, per se, location of residence with its concomitant implications for housing, educational, safety, environmental, and food access factors may also

have an impact on life expectancy (Robert Wood Johnson Foundation 2019). Variances in neighborhood outcome prevalence and resultant spatial patterns also reveal population health disparities that have been shaped by broader historical, social, political, and environmental issues which vary by location. Increasing digitization of health facility information, such as that from hospitals, clinics, and pharmacies, enables more in-depth analytics to improve public access, resource allocation, and service operation of healthcare. For example, NYC, Boston, and Chicago publish the geolocation of hospitals and clinics (NYC Department of Health and Mental Hygiene 2019a; Boston Department of Innovation and Technology 2019; Chicago Department of Public Health 2019). In addition, NYC also shares the geolocation of health facilities for purposes of hepatitis prevention and testing (NYC Department of Health and Mental Hygiene 2019b), and seasonal flu vaccination (NYC Department of Health and Mental Hygiene 2019c). Singapore maintains data on retail pharmacy locations (Singapore Health Sciences Authority 2019) as well as daily polyclinic attendance as categorized by selected diseases (Singapore Ministry of Health 2019). To fully understand current resource allocation, operations, and potential improvements, extensive analytical work, involving data mining and spatial analysis integrating with other population and environmental data, is required.

New survey approaches and data sources currently provide more granular information on behavioral factors related to urban health. City-level health surveys enable multifaceted research on health-related behaviors (e.g., smoking, drinking, physical activity, commuting patterns), as well as related household characteristics (e.g., age, gender, household size, foreign-born population), and socioeconomic status (e.g., income, education, occupation). Novel analytics using geotagged social media data make it possible to quantify, visualize, and promote public awareness of issues such as obesity, diabetes, or a physically active (or not) lifestyle (Ghosh and Guha 2013; Maitland et al. 2006; Hawn 2009). Analytics addressed to citizen complaints provide new insights into the relationship between heavy drinking and alcohol store location (Ransome et al. 2018), neighborhood risk from hazardous chemicals exposure (Gunn et al. 2017), and noise pollution varying by location and time (Zheng et al. 2014). Sensing technology and the IoT make it feasible to monitor environmental conditions at various spatial-temporal resolutions: At macro-mesoscale, satellite imagery and remote sensing data enable large scale spatial analysis on the health impact of land cover, ecological patterns, and natural disasters. At micro and hyperlocal scales, in situ sensing and GPS-enabled spatial tagging devices make it possible to monitor issues such as Dengue cases (Seidahmed et al. 2018), real-time ambient air quality (Schneider et al. 2017; Zheng et al. 2013) and drinking water quality (Hou et al. 2013).

21.3 Project Narrative: NYC PollenScape

21.3.1 Background

In 2015, more than 2,240 New Yorkers participated in *TreesCount! 2015*, a project hosted by the City Department of Parks & Recreation (NYC Department of Parks & Recreation 2015a). Each volunteer participant was issued a GPS device, a tape measure, and a training book to guide the digitization of any street tree's information, such as its geolocation (in latitude and longitude), species, size, health condition (e.g., damaged, overgrown, or dead), and sidewalk conditions (NYC Department of Parks and Recreation 2015b). To date, this is the largest crowd-sourced urban forestry data collection in the U.S. history (NYC Department of Parks & Recreation 2015c). The final output was reported in the NYC 2015 Street Tree Census which included data on more than 666,134 trees covering the streets in all five boroughs of NYC.

The research project *NYC PollenScape*¹ derived conceptually from some controversial findings on the health impact of urban trees. Generally, street trees function to clean the air (McPhearson et al. 2013), ease the urban heat island effect (Loughner et al. 2012), mitigate stormwater (Nowak et al. 2007), and create a more sustainable and aesthetic neighborhood that promotes physical activities (Ulmer et al. 2016). On the other hand, the potential adverse health impact of urban forestry has raised researchers' attention. Previous studies in the U.S. and Canada reveal an increasing health risk caused by tree pollen allergens. Certain tree species can be a source of allergens that exacerbate respiratory health issues including asthma (Lovasi et al. 2013). Surveys and clinical visit records reveal an underlying spatial-temporal correlation between allergenic pollen exposure and neighborhood asthma prevalence (Dales et al. 2008). Researchers in spatial epidemiology have further concluded that the local risk of pollen exposure should be incorporated in allergy diagnosis (Asam et al. 2015). Unfortunately, due to limited data sources and non-robust analytical methods, previous studies were constrained to specific case studies or small survey samples. More importantly, a lack of cross-domain knowledge integration and multi-disciplinary research have yielded inconsistent findings and implementations that are separately segmented within public health, environmental science, landscape architecture, and urban planning.

By 2016, the NYC Open Data platform already had over 1,600 related data sets publicly available (NYC Department of Information Technology 2016). These resources inspired us to consider additional street trees beyond the Department of Parks and Recreation's regular duties, with more integrated views on quality of life involving sustainability, safety, health, municipal services and beyond. Utilizing the tree census data and other ancillary data sets, project NYC PollenScape measures the localized environmental health impact of 600,000 + street

---

¹Full Publication: The impact of urban street tree species on air quality and respiratory illness: A spatial analysis of large-scale, high-resolution urban data: <https://www.sciencedirect.com/science/article/pii/S135382921830621X>.

trees of more than 120 species in NYC. Ultimately, this study aims to integrate the segmented data sets from various sectors representing the holistic urban physical-technical-ecological-socioeconomic dynamics that shape neighborhood respiratory health.

21.3.2 *Methodology*

Data mining and integration processes were employed to collect information from federal, state, and city sources, including the American Community Survey of the U.S. Census Bureau; the neighborhood asthma prevalence as captured by the New York State Department of Health; the citywide air quality monitoring program from the NYC Department of Health & Mental Hygiene; building and tax lot information from the NYC Department of City Planning; citizen complaints related to indoor air quality (e.g., chemical vapors, dry cleaning, construction dust) collected by NYC 311;² and public housing location as published by the NYC Housing Authority. Based on tree species, a web crawler extracts pollen information from Pollen.com including pollen allergenicity, severity, and active seasons. This integration creates an extensive database reporting—health outcomes (by asthma hospitalization rate), environmental conditions (e.g., street trees, pollen exposure, ambient air quality, indoor air quality), neighborhood demographics (population, age, income), and ‘built’ environment characteristics (building density, land use types, street network, public housing). These multiscale variables can be summarized at geolocation (in latitude and longitude), zip code, Census Block, or Neighborhood Tabulation Area scales through aggregation, disaggregation, spatial-interpolation, or spatial-extrapolation.

Asthma prevalence patterns are a result of complex biophysical-socioeconomic processes. Our modeling efforts aim to (1) specify those variables that capture underlying interactions among various factors and multicollinearity, and (2) estimate related interactions varying across space. The final model is a multivariate, geographically weighted regression model (GWR) which captures both a global coefficient β and a localized effect $\beta(ui, vi)$ that may vary by location (ui, vi) . A project website publishes the final results through interactive maps and plots built in Tableau.³ The general public can navigate the maps to check pollen exposure in their neighborhood during different seasons. A location-based spatial search engine enables the user, if in NYC, to zoom in on the neighborhood scale based on current location.

---

²A non-emergency service request hotline in New York City: <https://www1.nyc.gov/311/>.

³Project website: <https://urbanintelligencelab.github.io/NYCPollenScape/>.

21.3.3 Findings

The study revealed that the top 20 species represent more than 80% of street trees in NYC. Among these, many produce moderate or severely allergenic pollen. The peak pollen season is spring with 76% of street trees having active pollen during this period. The concentration of allergenic pollen exposure shifts by season, e.g. the South Bronx actually has a higher exposure risk during the fall. The regression models show that although street trees contribute to better air quality overall (measured by PM 2.5 concentrations), certain species (Red Maple, Northern Red Oak, and American Linden) are positively related to local asthma hospitalization rate (while other correlated factors are held fixed).

The GWR model results further explain the spatial disparities of environmental health in NYC which are collectively driven by ambient exposure, indoor environment quality, demographics, and socioeconomic status. For example, Midtown Manhattan has relatively high PM 2.5 concentrations but a lower asthma hospitalization risk than average; this might be explained by Midtown's higher income population having better awareness and access to preventive care as related to asthma. The spatial model also reveals a significant health risk associated with indoor air quality and public housing. Mott Haven is a low-income neighborhood in the south Bronx that has the city's highest youth asthma hospitalization rate. This aberrant rate is collectively driven by bad ambient air quality, pollen exposure, a vulnerable population, and poor housing quality (e.g., presence of 79% housing maintenance defects) (NYC Department of Health & Mental Hygiene 2015). These findings strongly suggest the need for cross-domain, intersectoral, multiscale data integration, and collaborative research efforts to understand urban health issues.

21.3.4 Limitations and Future Work

Long-term investigations and collaborative efforts are required to achieve integrated data intelligence for urban health as already noted. Admittedly, our current project has several limitations necessitating additional future work. Although new data sources enable comprehensive quantification of the urban context, neighborhood characteristics, and population baselines, a lack of granular health outcome data remains an important hurdle to developing data intelligence at high spatial-temporal resolution. To date, no publicly-available data reports specifically on asthma cases in NYC, at least in part because of privacy concerns and data ownership issues.

While it is promising to see more open data reporting regarding long term personal scale population health, there are some novel analytics which may serve as alternative approaches that can address the limitations noted. Since the direct data on asthma cases are often not available for *NYC PollenScape*, secondary analysis of other records may provide a proxy 'digital trace' of asthma patients. For example, Sheffield et al. analyzed 5-year asthma medication sales records in NYC and

found a significant correlation between pollen season and medication sales volume (Sheffield et al. 2003). With increasing deployment of the IoT in the urban environment, in situ sensing can potentially be employed to collect real-time or near real-time measurements as ground-truth validation. Currently, there are market-available sensors for monitoring particle concentrations (e.g., pollen, dust), air quality (e.g., PM_{2.5}, Ozone), and weather conditions (e.g., temperature, wind, precipitation). Our research may inform future in situ sensing at specific locations to better understand the complex interactions among air quality, micro-climate, pollen exposure, and asthma risk.

Considering that the NYC Tree Census data were collected through crowdsourcing, new civic analytics products may serve as an interface for (1) providing individuals with useful insights related to urban life as a return for their volunteering data collection efforts; and (2) collecting new information such as anonymous geotagging of user's neighborhood, capturing related urban environmental exposures, and the accrual of additional population health data. Long-term research testing on how to sustain a robust information feedback mechanism among city agencies, researchers, and citizens will also be necessary to achieve these goals.

21.4 Sociotechnical Considerations

Integrated urban data intelligence involves the environment, technology, and people. Cities as complex biophysical-technical-socioeconomic systems often raise challenges for developing technically feasible and socially viable solutions. Overall, successful research and deployment require careful consideration and understanding in order to address current and anticipated technical, social, and managerial challenges. Previous ad hoc "smart cities" deployments, legacy infrastructure, and enterprise-specific software applications have created a segmented data landscape in the urban domain (Harrison and Donnelly 2011). Data-driven decision-making and operations should respect specific urban contexts that may vary by historical, political, cultural, and regulatory environment factors. However, a lack of precise, transparent, and validated methodology across cities constrains more open and collaborative analytics. Data format, naming convention, and spatial unit definitions vary at different administrative scales including city, borough, community district, census tract, census block, and neighborhood tabulation area.

Methodological clarity becomes vital for developing reproducible, generalizable, and scalable computational solutions and analytical pipelines. In April 2017, the University of Chicago hosted the first workshop addressed to these issues entitled 'Convening on Urban Data Science' which included 112 experts from governmental agencies, universities, and the private sector. Presentations, discussions, and debates concluded with a common concern that the consistent analytical framework needed to address the rapid growth of data was not yet available.⁴ In addition, the advent

---

⁴Visit: <http://www.urbancced.org/urbandataconven>.

of artificial intelligence into the field was addressed: Algorithmic decision-making raises questions on how ‘black-box’ machine learning approaches can reliably handle underlying relationships (including the knotty issue of correlation versus causality) and confounding effects, while also taking into consideration the existing problematic urban patterns involving segregation and environmental justice.

Both ethics and social awareness must be addressed in carrying out data computing, analytics, and deployments that impact people’s lives. The ethical practice of data mining and analytics, especially regarding security and privacy issues, is critical for developing accountable methods, fair algorithms, and healthy partnerships with city agencies, stakeholders, and local communities (Bloomberg Data for Good Exchange 2017). Urban data is not always the ground-truth due to limited representativeness (e.g., survey data), reporting biases (compliant data), or the performative nature of specific behavior (social media data). Hence, data scientists need to be fully aware of the limitations of specific data sources or types. Decisions that involve physical infrastructure, capital investment, and policy intervention are often irreversible within the short-term, while an A/B testing is neither feasible nor ethical in reality. Data scientists should work with policy-makers and planners to carefully evaluate potential risks.

Managerial and domain barriers constrain multidisciplinary research and practice in cities. Urban health issues involve various biophysical and socioeconomic factors that require cross-domain efforts and intersectoral actions (World Health Organization 2008). In reality, cities are complex systems-of-systems with agencies often operating within a silo, creating managerial and organizational barriers. Although city agencies are becoming data-rich, different departmental demands and operations often come to define data collection, analytics, and management. Integrated analytics face real-life constraints shaped by administrative hierarchies, organizational priorities, and competing interests. Besides the managerial silos, collaborative urban analytics needs to break down the domain barriers.

Multifaceted urban issues require trans-disciplinary approaches integrating science, engineering, and design expertise to address both social and technical urban problems. In 2016, the NYC Department of Parks and Recreation organized *TreeCount! Data Jam*,⁵ a one-day hackathon that invited the general public to explore potential insights from the street tree census data (NYC Department of Parks and Recreation 2016). During this event, urban planners, tree enthusiasts, data scientists, and students from universities formed teams to analyze data, develop research questions, design prototypes, and visualize the potential use cases. However, the effective facilitation and maintenance of robust multidisciplinary collaborative research are on-going challenges.

Effective and sustained partnership is a crucial enabler for the information feedback loops needed to support successful long-term implementations. Since cities are complex systems-of-systems, their overall success relies on great efforts that are required for integration, communication, and engagement (Maier 1998). A regional cross-cities network enables information exchange and experience

---

⁵Visit: <https://beta.nyc/2016/05/20/nyc-treescount-data-jam-challenges/>.

sharing. For example, MetroLab Network is a U.S. city-university league created to promote civic technology.⁶ Within cities, collaboration opportunities also lie at the policy-academic-industry nexus. In NYC, the Mayor's Office of the Chief Technology Officer launched 'NYCx Challenges' to support local business partners and researchers on pilot projects promoting sustainability, health, and economic development.⁷ City-university-community partnerships enable developing "test-beds" to explore how information technology and data science may improve the quality of life at a neighborhood scale. Research/community innovation projects, such as the 'Array of Things' project by the University of Chicago⁸ and the 'Quantified Community' project by New York University,⁹ provide first-hand experience in innovative technology deployment, data-driven operations, and citizen sciences.

21.5 Conclusion

The rapid digitization of urban life brings opportunities for developing new methods and applications to promote urban health. As we live in increasingly smart and connected society, cross-domain integration and participatory research play important roles in addressing previous ad hoc technology development, and top-down urban policies and operations. Citizen-involved and community-based projects will continue (1) utilizing various technologies to solve neighborhood problems, and local demands for better quality of life, (2) educating the public for better data literacy and promoting citizen science through active engagement, and (3) validating the progress and impact of urban digital transformation at a granular human scale. All of these technical, methodological, and social transformations working progressively in tandem will result in the creation of a new data-driven version of urban science.

References

- The U.S. Environmental Protection Agency. (2019). Airnow api. <https://docs.airnowapi.org>.
 Asam, C., Hofer, H., Wolf, M., Aglas, L., & Wallner, M. (2015). Tree pollen allergens-an update from a molecular perspective. *Allergy*, 70, 1201–1211.
 Barbosa, L., Pham, K., Silva, C., Vieira, M. R., & Freire, J. (2014). Structured open urban data: Understanding the landscape. *Big data*, 2, 144–154.
 Bloomberg Data for Good Exchange. (2017). Bloomberg, brighthouse and data for democracy launch initiative to develop data science code of ethics. <https://www.bloomberg.com/company/announcements/bloomberg-brighthouse-data-democracy-launch-initiative-develop-data-science-code-ethics/>.

---

⁶<https://metrolabnetwork.org>.

⁷<http://www.nyc.gov/html/nycx/challenges.html>.

⁸<https://arrayofthings.github.io>.

⁹<https://www.fastcompany.com/3029255/beyond-the-quantified-self-the-worlds-largest-quantified-community>.

- Boston Department of Innovation and Technology. (2019). Hospital locations. <https://data.boston.gov/dataset/hospital-locations>.
- Bradley, G. (2007). *Social and community informatics: Humans on the net*. Routledge
- Chicago Department of Public Health. (2019). Clinic locations. <https://data.cityofchicago.org/Health-Human-Services/Chicago-Department-of-Public-Health-Clinic-Locations/kcki-hnch>.
- Dales, R. E., Cakmak, S., Judek, S., & Coates, F. (2008). Tree pollen and hospitalization for asthma in urban Canada. *International Archives of Allergy and Immunology*, 146, 241–247.
- FitBit. (2019). Fitbit api. <https://dev.fitbit.com>.
- Forrester, J. W. (1970). Systems analysis as a tool for urban planning. *IEEE Transactions on Systems Science and Cybernetics*, 6, 258–265.
- Ghosh, D., & Guha, R. (2013). What are we tweeting about obesity? mapping tweets with topic modeling and geographic information system. *Cartography and geographic information science*, 40, 90–102.
- Gunn, L. D., Greenham, B., Davern, M., Mavoa, S., Taylor, E. J., & Bannister, M. (2017). Environmental justice in Australia: Measuring the relationship between industrial odor exposure and community disadvantage. In *Community quality-of-life indicators: Best cases VII* (pp. 113–133). Springer.
- Harrison, C., & Donnelly, I. A. (2011). A theory of smart cities. In *Proceedings of the 55th Annual Meeting of the International Society for the Systems Sciences* (pp. 55).
- Hawn, C. (2009). Take two aspirin and tweet me in the morning: how twitter, facebook, and other social media are reshaping health care. *Health Affairs*, 28, 361–368.
- Hou, D., Song, X., Zhang, G., Zhang, H., & Loaiciga, H. (2013). An early warning and control system for urban, drinking water quality protection: Chinas experience. *Environmental Science and Pollution Research*, 20, 4496–4508.
- James, M., Michael, C., Brad, B., Jacques, B., & Richard, D., Charles, R., et al. (2011) Big data: The next frontier for innovation, competition, and productivity, Technical Report, McKinsey & Company. <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>.
- Loughner, C. P., Allen, D. J., Zhang, D.-L., Pickering, K. E., Dickerson, R. R., & Landry, L. (2012). Roles of urban tree canopy and buildings in urban heat island effects: Parameterization and preliminary results. *Journal of Applied Meteorology and Climatology*, 51, 1775–1793.
- Lovasi, G. S., O’Neil-Dunne, J. P., Lu, J. W., Sheehan, D., Perzanowski, M. S., MacFaden, S. W., et al. (2013). Urban tree canopy and asthma, wheeze, rhinitis, and allergic sensitization to tree pollen in a New York City birth cohort. *Environmental Health Perspectives*, 121(2013), 494.
- Maier, M. W. (1998). Architecting principles for systems-of-systems. *Systems Engineering: The Journal of the International Council on Systems Engineering*, 1, 267–284.
- Maitland, J., Sherwood, S., Barkhuus, L., Anderson, I., Hall, M., & Brown, B., et al. (2006). Increasing the awareness of daily activity levels with pervasive computing, In *2006 Pervasive Health Conference and Workshops* (pp. 1–9). IEEE.
- McPhearson, T., Maddox, D., Gunther, B., & Bragdon, D. (2013). Local assessment of New York City: Biodiversity, green space, and ecosystem services. In *Urbanization, biodiversity and ecosystem services: Challenges and opportunities* (pp. 355–383). Springer.
- National Association for Public Health Statistics and Information Systems (NAPHSIS). (2018). United states small-area life expectancy project (USALEEP). <https://www.naphsis.org/usaleep>.
- Nowak, D. J., Robert III, E., Crane, D. E., Stevens, J. C., & Walton, J. T. (2007). Assessing urban forest effects and values, New York City’s urban forest, Technical Report, United States Department of Agriculture, Forest Service, Northern Research Station. <https://www.nrs.fs.fed.us/pubs/rb/rbns009.pdf>.
- NYC Department of Health & Mental Hygiene. (2015). New York City Community Health Profiles. <https://www1.nyc.gov/assets/doh/downloads/pdf/data/2015chp-bx1.pdf>.
- NYC Department of Health and Mental Hygiene. (2018). NYC community health profiles. https://www1.nyc.gov/site/doh/data/data-publications/profiles.page#bx_9.

- NYC Department of Health and Mental Hygiene. (2019a). New York City hospitals. <https://data.cityofnewyork.us/Health/hospital/q6fj-vxf8>.
- NYC Department of Health and Mental Hygiene. (2019b). Dohmh health map - hepatitis. <https://data.cityofnewyork.us/Health/DOHMH-Health-Map-Hepatitis/nk7g-qeep/data>.
- NYC Department of Health and Mental Hygiene. (2019c). New York City locations providing seasonal flu vaccinations. <https://data.cityofnewyork.us/Health/New-York-City-Locations-Providing-Seasonal-Flu-Vac/w9ei-idxz/data>.
- NYC Department of Information Technology & Telecommunications. (2016). NYC open data dashboard. <https://opendata.cityofnewyork.us/dashboard/>.
- NYC Department of Parks & Recreation. (2015a). 2015 Street tree census report. <http://media.nycgovparks.org/images/web/TreesCount/Index.html>.
- NYC Department of Parks and Recreation. (2015b). 2015 Street tree census report. <http://media.nycgovparks.org/images/web/TreesCount/Index.html>.
- NYC Department of Parks & Recreation. (2015c). Treescount!2015. <https://www.nycgovparks.org/trees/treescount>.
- NYC Department of Parks and Recreation. (2016). Treescount! data jam. <https://www.nycgovparks.org/events/2016/06/04/treescount-data-jam>.
- Ransome, Y., Luan, H., Shi, X., Duncan, D. T., & Subramanian, S. (2018). Alcohol outlet density and area-level heavy drinking are independent risk factors for higher alcohol-related complaints. *Journal of urban health* 1–13.
- Robert Wood Johnson Foundation. (2019). Could where you live influence how long you live? <https://www.rwjf.org/en/library/interactives/whereyouliveaffectshowlongyoulive.html>.
- Schneider, P., Castell, N., Vogt, M., Dauge, F. R., Lahoz, W. A., & Bartonova, A. (2017). Mapping urban air quality in near real-time using observations from low-cost sensors and model information. *Environment International*, 106, 234–247.
- Seidahmed, O. M., Lu, D., Chong, C. S., Ng, L. C., & Eltahir, E. A. (2018). Patterns of urban housing shape dengue distribution in Singapore at neighborhood and country scales. *GeoHealth*, 2, 54–67.
- Sheffield, P. E., Weinberger, K. R., Ito, K., Matte, T. D., Mathes, R. W., Robinson, G. S., et al. (2011). The association of tree pollen concentration peaks and allergy medication sales in New York City: 2003–2008. *ISRN Allergy* 2011.
- Singapore Health Sciences Authority. (2019). Retail pharmacy locations. <https://data.gov.sg/dataset/retail-pharmacy-locations>.
- Singapore Ministry of Health. (2019). Average daily polyclinic attendances for selected diseases. <https://data.gov.sg/dataset/average-daily-polyclinic-attendances-selected-diseases?viewid=8fb8637d-c1c5-4c5e-9fbe-3f46785804b7&resourceid=dd4dcaac-aa8d-49de-a96a-b809f8d3ae0d>.
- Twitter. (2019). Twitter for developer. <https://developer.twitter.com/en.html>.
- Ulmer, J. M., Wolf, K. L., Backman, D. R., Tretheway, R. L., Blain, C. J., O’Neil-Dunne, J. P., et al. (2016). Multiple health benefits of urban tree canopy: The mounting evidence for a green prescription. *Health & Place*, 42, 54–62.
- U.S. Census Bureau. (2010). Decennial census. <https://www.census.gov/history/www/programs/demographic/decennialcensus.html>.
- U.S. Census Bureau. (2018). American community survey (ACS). <https://www.census.gov/programs-surveys/acs/about.html>.
- U.S. Department of Health and Human Services. (2019). Aidsinfo api. <https://aidsinfo.nih.gov/api>.
- Von Bertalanffy, L. (1969). General systems theory and psychiatry-an overview. *General Systems Theory and Psychiatry*, 32, 33–46.
- Wiener, N. (1948). Cybernetics. *Scientific American*, 179, 14–19.
- Wiig, A. (2015). IBM’s smart city as techno-utopian policy mobility. *City*, 19, 258–273.
- Wolf, G. (2007). Know thyself: Tracking every facet of life, from sleep to mood to pain, 24/7/365, *Wired Magazine*.
- World Health Organization. (2008). Intersectoral action. <https://www.who.int/socialdeterminants/thecommission/countrywork/within/isa/en/>.

- Zheng, Y., Liu, F., & Hsieh, H.-P. (2013). U-air: When urban air quality inference meets big data. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1436–1444). ACM.
- Zheng, Y., Liu, T., Wang, Y., Zhu, Y., Liu, Y., & Chang, E. (2014). Diagnosing new york city's noises with ubiquitous data. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 715–725). ACM.
- Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*. Addison-Wesley press.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.