Hyper-local Urban Contextual Awareness through Open Data Integration

Yuan Lai
Department of Urban Studies and Planning
Massachusetts Institute of Technology
Cambridge, USA
yuanlai@mit.edu

***Abstract*—Increasing data generated in cities provide new opportunities for analyzing urban characteristics at high resolution. Data integration can enhance our understanding in urban systems and related operations for economic growth, social equity, and environmental sustainability. This article provides a comprehensive overview of urban data collection in New York City and proposes a method for hyper-local open data mining and integration. We first introduce the research context, data sources, and major analytical approaches that are generalizable for any given location. We then apply the method to all street intersections in Manhattan (n=3,077) to demonstrate potential implementations of hyper-local intelligence for data-driven decision-making in cities. The paper concludes with a discussion on limitations and future work.**

***Index Terms*—urban informatics, urban science**

I. Introduction

The increasing availability of data from city agencies, commercial transactions, social media, and sensors provides new information sources for analyzing urban dynamics at hyper-local resolution. However, such data often exist at various spatial units and temporal frequencies, exhibit a range of data quality and standardization problems caused by agencies and sectors operating in siloes, guided by specific needs. This fragmentation constrains a holistic understanding of urban dynamics across multiple urban systems and sub-systems, characterized by social, economic, physical, and environmental dimensions. Data integration and knowledge mining play key roles in transforming data into actionable insights that can be used to improve city operational, policy, and planning decisions. Location-based information plays a key role in cities since planning decisions, design solutions, and service operations are often driven and defined by place [1]. Increasing volume, variety, and velocity of urban data not only extends our view of cities as large complex systems but also can be used to advance our fundamental understanding of the dynamics of place at high spatial resolution. 'Hyper-local' is a geographic scale smaller than any current existing spatial boundary, such as census block, Zip code, or neighborhood [2]. A hyper-local area is a small region based on a given geo-location or a cluster of geo-tagged data points [3].

This study focuses on hyper-local knowledge mining as a process of integrating data, extracting information, and generating purposeful insights based on a geo-location surrounding physical, ecological, cultural, and socioeconomic characteristics, known as hyper-local urban contextual awareness. Previous studies have defined knowledge in this context as 'a combination of data, information, and analytics associated with human cognition' [4]. One of our previous studies proposed a data-driven method to quantify hyper-local urban characteristics through data integration and spatial query [5]. Using 100 locations in New York City (NYC), our machine learning-based classifier identifies three types of places: (1) job and transit-oriented places, (2) residential area with low density, and (3) mixed-use area with relative better streetscape measured by local street trees, public amenities, sidewalk area, and pavement condition. The findings support a generalizable method for extracting multi-sources open data to create high-dimensional measures of urban life and location-based characteristics. However, previous study suffers from limited locations with questions regarding data representativeness and sampling biases.

This study extends the exploration in hyper-local data mining and urban typology classification by quantifying almost every street intersection in Manhattan. The purposes are threefold: First, we re-exam open data mining and integration on a much larger scale (n=3077) to test generalizability. Second, we investigate new questions building upon previous findings on how urban typology classifier performs with the increasing number of places. Third, we explore the potential value of hyper-local intelligence with every Manhattan's street intersection quantified and classified.

This paper articulates the motivation, data resources, and a new methodology for mining hyper-local knowledge through open data integration in cities. We begin with a literature review on the definition and context of location-based knowledge, as well as its value in urban planning, service operation, and business intelligence. Using Manhattan as a case study, we introduce major data sources and the context of data collection and management. We then summarize major approaches for data integration and knowledge extraction at high spatial resolutions. Finally, we propose several potential use cases of hyper-local urban intelligence, particularly for the public good. The paper concludes with a discussion on potential applications, limitations, and future work.

II. Data & Methods

Urban Data Sources

Urban open data[1] reflect biophysical form, socioeconomic activities, logistical operations, and civic engagement in cities, among other variables [7], [8]. Table I summarizes selected urban datasets that are available in NYC. Government public data are curated and published by the city, state, and federal agencies to promote transparent governance, as well as digital entrepreneurship and civic engagement [7], [8]. In 2012, NYC passed Local Law 11 known as 'Open Data Law' that requires city agencies to make administrative data publicly available and accessible through a common digital portal known as NYC Open Data [9]. Agencies report data on different subsystems of the city, including buildings, land use and development, transportation, public space, and environmental quality [10], among others. Each agency collects, manages, and publishes data of primarily three types. The first type includes the digitization of urban physical systems, including land use, buildings, street network, street trees, and transit facilities. These datasets represent the physical components in the urban built environment that are relatively static with batch updates at some regular frequency. The Department of City Planning publishes Primary Land Use Tax Lot Output (PLUTO), an annually-updated dataset on land use and building characteristics at the tax lot level [11]. It includes data on land use type, built area, space usage, zoning, the total number of units, and land assessed value, extracted with a spatially-computable geometry identified by a unique identifier known as the Borough-Block-Lot number. The second type represents 'naturally occurring' data as the digital exhaust of operational or transactional activities, such as digital records on building violations, property sale transactions, construction permit applications, and 311 service requests [7]. Such record data usually either report by a unique identifier matched with a particular data inventory or as a geo-point in latitude and longitude coordinates. The third type includes survey data by federal, state, and city agencies, including population census data, neighborhood health surveys, pedestrian counts, or environmental quality surveys. Such data are largely used in conventional quantitative urban research, but with limitations created by the often low resolution, incomplete coverage, small sample size, and high cost of collection. LEHD[2] reports census block level population by origin-destination employment statistics by the US Census Bureau [12]. This data provides information on local population composition, as well as underlying commuting patterns based on work-home locations.

Rapid development of pervasive computing generates large volume and variety of sensing data, which enable the instrumentation of urban systems [13]. Intentional sensing utilizes sensors to subjectively collect data, such as conventional in-situ technical sensing using sensors and microprocessors, remote sensing by satellites, and human sensing using infrared counters [14], [15]. Unintentional sensing utilizes 'by-product' data generated by human or machine systems. Applications include mapping human mobility through cellular and local WiFi network usage, evaluating built space using imagery data, or tracking public sentiment using social media data [16]–[21]. Recent studies utilize computer vision to measure the visual appearance of urban space and further analyze the psychological and behavioral impact of built environment [20], [22]–[24]. Recent interests in citizen science promote participatory sensing that engages local communities to monitor quality-of-life factors such as air quality and noise [25], [26].

Increasing private enterprises start to compete or disrupt traditional urban systems in media, transportation, housing, or commerce [27]. This transition makes private enterprises as an alternative urban data provider. Many data-intensive companies such as Twitter, Zillow and Yelp regularly make partial data publicly available through APIs to encourage open source development and academic research. Other enterprises have built digital products that provide customers timely information as business intelligence. For example, Mastercard provides a digital platform visualizing time-series of total monthly spending at ZIP code level aggregated by purchasing categories [28]. These digital tools enable business owners to monitor local trends, identify customer groups, and their collective preferences.

Data Integration, Quantification, and Classification

Urban open data, however, derive from different sources with various spatial resolution, temporal frequency, and data quality. A fundamental challenge in quantifying places is how to generate meaningful and comparable measures using data mining and information integration techniques. Recent studies highlight increasing interests in hyper-local knowledge mining by integrating cross-domain datasets [14], [29], [30]. Data extraction characterizes a hyper-local area by extracting urban level data [14]. It aims to explain a particular local phenomenon with surrounding geospatial attributes or variances across locations. Computationally, this process estimates a given location by aggregating a series of citywide datasets using geometric intersects, spatial join, or table merge [31]. For a geo-location, a general approach is to generate circular buffers or convex hulls to extract data on land use, buildings, street networks, and facilities [32]–[35].

We develop a spatial query algorithm to extract quantitative measures based on street intersections (n=3077) identified from NYC Single Line Street Base Map. Using each node, it generates a 1/8 mile radius circular buffer and intersects with multiple datasets. For location-based point data, such as trees, public amenities (public benches, bike racks, digital kiosks), and transit connections (bus stops, subway stations, and bike-share stations), a buffer can check availability as a binary variable and total count as a numerical variable. For data in different spatial resolution defined by administrative boundary and neighborhood, we use areal interpolation to integrate spatial data at a unified spatial scale [36], [37]. We estimate the hyper-local worker/ resident population using

[1] 'Open Data' are datasets publicly available for use and distribution without restrictions regarding privacy, confidentiality, or security concerns [6].

[2] Longitudinal Employer-Household Dynamics by US Census Bureau.

TABLE I
SELECTED NYC URBAN DATA SOURCES

Category	Data	Resolution	Frequency
Population	US Census	Census Block	Decennial
	American Community Survey	Neighborhood Tabulation Area	
	Longitudinal Employer-Household Dynamics	Census Block	Annual
Land, Topography & Buildings	Primary Land Use Tax Lot Output	Building	Annual
	LiDAR	Point	NA
	Construction permits	Tax lot	Daily
	Violations	Building	Daily
	Energy& water consumption	Building	Annual
	Property sales	Unit	Daily
Public Space & Urban Design	Parks properties	Land parcel	NA
	Sidewalk area	Shapefile	NA
	Street bench	Point	NA
	Digital kiosks	Point	NA
	OpenStreetMap	Polyline	NA
	Street trees	Point	Decennial
	LION Single Line Street Base Map	Segment	NA
	Street score	Points	NA
Transportation & Public Transit	Pedestrians Count	Point	Bi-annual
	Average Daily Traffic	Point	Annual
	Subway station	Point	NA
	Subway turnstile counts	Point	Hourly
	Bus stops	Point	NA
	Travel survey	Household	Annual
	MTA vehicle location	Point	Timestamp
	Taxi pick-ups & drop-offs	Point	Timestamp
Operation & Civic Services	Air Quality Monitoring	United Hospital Fund Area	Annual
	311 service request	Point	Timestamp
	NYPD major crime		
	Waste collection		
	Emergency response		
Data-driven Business	Zillow sales data	ZIP code	Monthly
	Citibike bike-share system data	Point	Timestamp
	Twitter feeds		
	Waze traffic condition feeds		
	Google Street Views	Image	NA

TABLE II
HYPER-LOCAL URBAN CONTEXT METRIC QUATIFICATION

Metric	Description
Density	Building density
Home	Percentage of residents in total local population
Streetscape	Density of street trees and public amenities
Retail	Percentage of retail space
Transit	Accessibility to subway and buses
Jobs	Total number of employees
Space Diversity	An entropy index of land use mix
Income	Neighborhood median household income

LEHD weighted by land use characteristics from PLUTO. The local population can be estimated by integrating multiple datasets at census tract, census block, and tax lot level.

The quantification process combines hyper-local data extractions into quantitative metrics that are meaningful for urban planning decision-making, operation, and public services. We identify eight critical metrics relevant to location-based services in cities, including building density, home, streetscape quality, retail, transit, job concentration, land use diversity, and neighborhood income (Table II). Normalization uses Z-normalization to standardize different metrics to be comparable with a mean of zero and a standard deviation of one. Agglomerative hierarchical clustering method can classify all intersections as urban typologies with eight normalized metrics. The dendrogram of hierarchical clustering indicates three as the optimal number of clusters (Figure 1). We use radar charts to plot multivariate hyper-local characteristics into two-dimensional visualization.

III. FINDINGS

The data integration process reveals that hyper-local contextual awareness requires high-dimensional data representations integrated from multiple sources. Previously, there are well-established techniques for analyzing a particular urban sub-system (e.g., transportation) that focuses on geometrical configuration of street network or built form [38]. However, such methods fall short when analyzing complex urban phenomena such as pedestrian behavior, health, land use and urban development, and social behavior in cities [39]. Quantification of building density, street networks, or tree coverage at the hyper-local scale represents the physical infrastructure of the

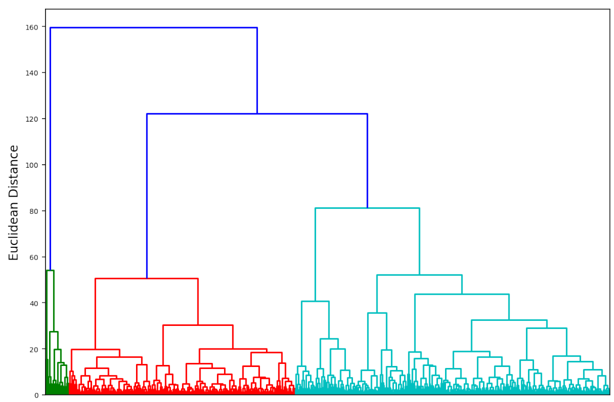

Fig. 1. Dendrogram of place classification using hierarchical clustering.

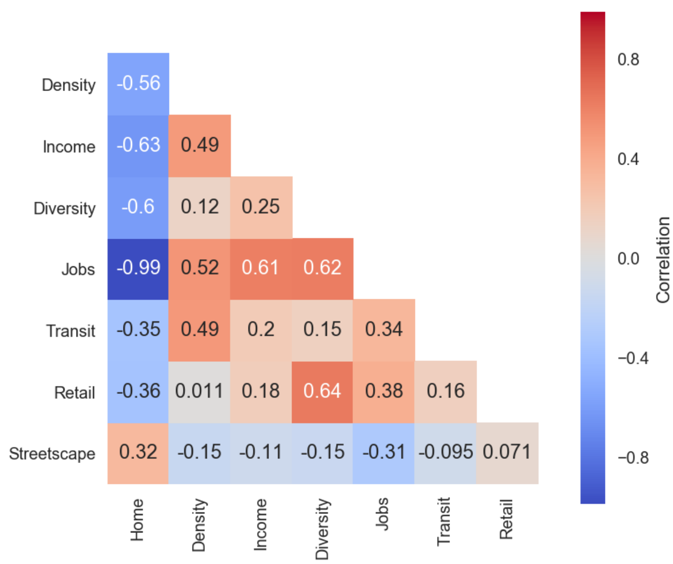

Fig. 2. Correlation matrix of urban context measures at hyper-local scale.

city that changes with relatively long temporal cycles. Such heterogeneity emphasizes the spatial variance within a single urban sub-system, with less consideration given to temporal dynamics or interaction across multiple sub-systems. A holistic understanding in places requires a generalizable quantification process integrating various data.

Figure 2 compares correlations among hyper-local measures, reflecting the loss and gain of characteristics by place. Urban design quality and people's perception of places require more integrated models that include built environment characteristics, points-of-interest, land topography, and socioeconomic factors [33], [40]. For example, 'Street Score', the data used for quantifying streetscape, derives from street view images and crowd-sourced feedback [20]. Our results indicate the perception of street quality may associate with the feeling of residential. Therefore, even though the Street Score is a relatively 'objective' measure based on public feedback, it requires careful interpretation with other local characteristics.

Classification results indicate three dominant types of places: (I) places with high density, concentration of jobs, mixed land use, charming streetscape, and transit connections; (II) extreme dense area with job concentration but a lack of land-use diversity and streetscape; and (III) low-density residential area but with limited job destinations, transit, and commercial space (Figure 3). Figure 4 visualizes all quantified and classified street intersections in Manhattan. While increasing data sources extend understanding in urban spatial-temporal dynamics, they raise new challenges for calibrating and comparing observations across different locations [14], [15]. Our typology classification, along with quantified measures, provides contextual awareness for Internet-of-Things (IoT) and ubiquitous computing in cities [14]. By November 2019, NYC has installed about 1168 smart kiosks on streets in Manhattan[3]. Quantified local context and characteristics enable more location-based customization for such urban IoT devices, including place-specific information services and neighborhood engagement. NYC Department of Environmental Protection, as another example, conducts quarterly local air quality monitoring at 162 locations [41]. Again, our approach can quantify and classify any location to generate baseline measures for interpreting hyper-local air quality data. In this way, we can better understand micro-climate conditions and environmental quality in different types of places.

These findings are also relevant to data-driven decisions, public service operation, and location-based policy in cities. As planning and policy-making become increasingly data-driven, urban analytics gradually shifts from mathematical models or simulation into real-world data computing [15]. Planners have long-term interests in analyzing complex socioeconomic urban issues. Since a controlled experiment is often not feasible in an urban context, a data-driven approach may unlock our understandings in local phenomena through a collection of multidisciplinary factors. Big data, paired with advanced computing, may reveal complex urban system behaviors with anomalies, spatial hot-spots, or temporal patterns. Data science techniques can pinpoint the problem, track the source, or predict the progress of urban development at high resolution [42].

IV. Discussion & Conclusion

Big data, IoT, and machine learning accelerate the convergence among the physical, digital, and social layers of cities [43]. Distinctive from typical 'big data', the challenge of urban data is not solely the large volume, variety, and velocity, but fragmented, messy, and unstructured information [7]. The value of integrated hyper-local knowledge is still largely untapped, mainly due to a lack of method, technical difficulties, and ethical concerns. As city agencies and organizations develop in-house analytical tools for specific problems, methodological ambiguities rise with reproducibility, scalability, and accountability issues. For instance, there are at least five organizations in NYC investing different technologies and analytical methods for pedestrian counting, without common data

[3]LinkNYC: https://www.link.nyc

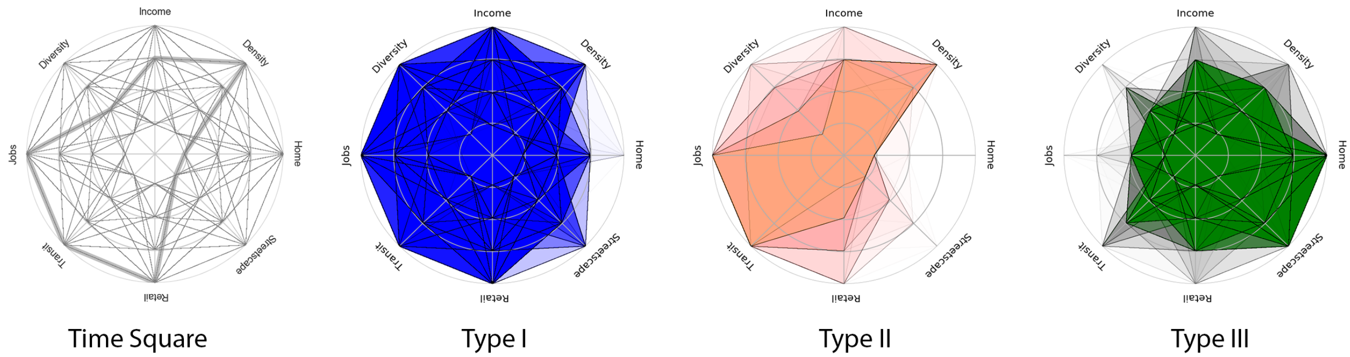

Fig. 3. Three types of places in Manhattan based classified by clustering algorithm with eight metrics based on real-world data.

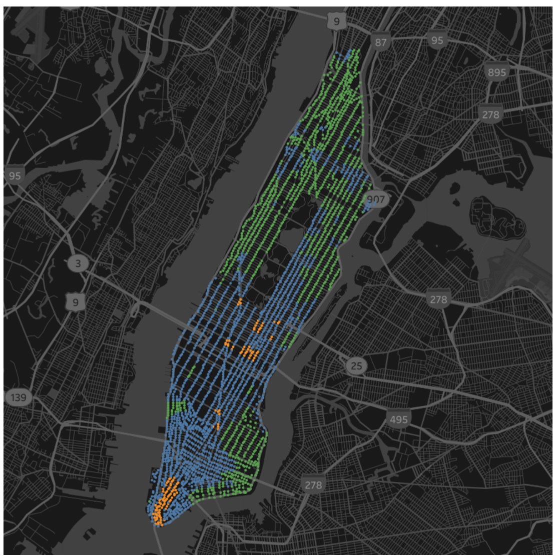

Fig. 4. Quantified and classified street intersections in Manhattan.

definition or quality standards [44]–[48]. This constrains data sharing and result validation across agencies and organizations and reduces confidence in mobility models derived from these data. Methodological ambiguities create a fragmented urban digital landscape that further brings technical difficulties for data integration. Since data are generated in siloes that are often not compatible across operating or database systems, integration is not a simple concatenation, but a process that necessitates both data science and urban domain expertise. Finally, hyper-local data provide information at high resolution, bringing ethical concerns related to privacy, security, and data ownership to the forefront. Since data analytics in the urban context is intended to drive operational and policy decisions, particular attention must be given to issues of bias and the trade-off between privacy and granular understandings.

There are certainly limitations in open data integration regarding the nature of both urban systems and human activity. Hyper-local knowledge mining may provide interesting facts to residents at microscale, but how it can improve long-term planning or policy is mostly undefined. More granular, diverse, and frequent data increase model complexity, but it is uncertain how they can provide actionable insights for operation rather than nuances. Representativeness and accountability of new information sources are still questionable. Social media data, for instance, are communicative and performative by nature of the activity, so they may not represent the full picture of urban mobility and citizen activities [42].

Future research interest is two-fold. First, we are interested in integrating hyper-local urban contextual information with real-time situational awareness by introducing high-frequency spatiotemporal data, such as emergency calls, citizen complaints, and geo-tagged social media. Another research interest is to test whether and how pre-computed hyper-local parameters can enhance machine intelligence built upon non-structural data from video, images, and audio.

In conclusion, increasing volume, granularity, and quality of open data create new opportunities for quantifying urban characteristics at a hyper-local scale. The value of urban data integration is still largely untapped, considering how it can contribute to engineering, planning, design, and operation in cities. The nature of urban data also raises technical, methodological, and ethical challenges that require proper integration techniques and analytical pipelines. Although cities flourish with increasing open data, accountability and effectiveness of hyper-local data integration still need to be further validated.

REFERENCES

[1] S. P. French, C. Barchers, and W. Zhang, "How should urban planners be trained to handle big data?" in *Seeing Cities Through Big Data*. Springer, 2017, pp. 209–217.

[2] C. Xia, R. Schwartz, K. Xie, A. Krebs, A. Langdon, J. Ting, and M. Naaman, "Citybeat: real-time social media visualization of hyper-local city data," in *Proceedings of the 23rd International Conference on World Wide Web*. ACM, 2014, pp. 167–170.

[3] D. Flatow, M. Naaman, K. E. Xie, Y. Volkovich, and Y. Kanza, "On the accuracy of hyper-local geotagging of social media content," in *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. ACM, 2015, pp. 127–136.

[4] C. Starkey and C. Garvin, "Knowledge from data in the built environment," *Annals of the New York Academy of Sciences*, vol. 1295, no. 1, pp. 1–9, 2013.

[5] Y. Lai and C. E. Kontokosta, "Quantifying place: Analyzing the drivers of pedestrian activity in dense urban environments," *Landscape and Urban Planning*, vol. 180, pp. 166–178, 2018.

[6] M. Janssen, Y. Charalabidis, and A. Zuiderwijk, "Benefits, adoption barriers and myths of open data and open government," *Information systems management*, vol. 29, no. 4, pp. 258–268, 2012.

[7] P. V. Thakuriah, N. Y. Tilahun, and M. Zellner, "Big data and urban informatics: innovations and challenges to urban planning and knowledge discovery," in *Seeing Cities Through Big Data*. Springer, 2017, pp. 11–45.

[8] R. Krishnamurthy, K. L. Smith, and K. C. Desouza, "Urban informatics: critical data and technology considerations," in *Seeing Cities Through Big Data*. Springer, 2017, pp. 163–188.

[9] NYC Department of Information Technology & Telecommunications. (2018) Open data law. [Online]. Available: https://www1.nyc.gov/site/doitt/initiatives/open-data-law.page

[10] The City of New York. (2018) NYC Open Data. [Online]. Available: https://opendata.cityofnewyork.us/

[11] PLUTO. (2016) PLUTO and MapPLUTO data. New York City Department of City Planning. [Online]. Available: https://www1.nyc.gov/site/planning/data-maps/open-data/dwn-pluto-mappluto.page

[12] LEHD. (2016) Longitudinal Employer-Household Dynamics Origin-Destination Employment Statistics (LODES). U.S. Census Bureau. [Online]. Available: https://lehd.ces.census.gov/data/

[13] J. Lambrechts and S. Sinha, *Microsensing Networks for Sustainable Cities*. Springer, 2016.

[14] G. Sagl, B. Resch, and T. Blaschke, "Contextual sensing: Integrating contextual information with human and technical geo-sensor information for smart cities," *Sensors*, vol. 15, no. 7, pp. 17 013–17 035, 2015.

[15] L.-M. Ang and K. P. Seng, "Big sensor data applications in urban environments," *Big Data Research*, vol. 4, pp. 1–12, 2016.

[16] S. Jiang, J. Ferreira, and M. C. Gonzalez, "Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore," *IEEE Transactions on Big Data*, vol. 3, no. 2, pp. 208–219, 2017.

[17] C. E. Kontokosta and N. Johnson, "Urban phenology: Toward a real-time census of the city using Wi-Fi data," *Computers, Environment and Urban Systems*, vol. 64, pp. 144–153, 2017.

[18] L. Yin, Q. Cheng, Z. Wang, and Z. Shao, "'Big data' for pedestrian volume: Exploring the use of google street view images for pedestrian counts," *Applied Geography*, vol. 63, pp. 337–345, 2015.

[19] A. G. Rundle, M. D. Bader, C. A. Richards, K. M. Neckerman, and J. O. Teitler, "Using Google Street View to audit neighborhood environments," *American journal of preventive medicine*, vol. 40, no. 1, pp. 94–100, 2011.

[20] N. Naik, J. Philipoom, R. Raskar, and C. Hidalgo, "Streetscore - predicting the perceived safety of one million streetscapes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014, pp. 779–785.

[21] C. Wu, X. Ye, F. Ren, Y. Wan, P. Ning, and Q. Du, "Spatial and social media data analytics of housing prices in shenzhen, china," *PloS one*, vol. 11, no. 10, p. e0164553, 2016.

[22] J. Zacharias, "Pedestrian behavior pedestrian behavior and perception in urban walking environments," *Journal of Planning Literature*, vol. 16, no. 1, pp. 3–18, 2001.

[23] P. Salesses, K. Schechtner, and C. A. Hidalgo, "The collaborative image of the city: mapping the inequality of urban perception," *PloS one*, vol. 8, no. 7, p. e68400, 2013.

[24] S. J. Mooney, C. J. DiMaggio, G. S. Lovasi, K. M. Neckerman, M. D. Bader, J. O. Teitler, D. M. Sheehan, D. W. Jack, and A. G. Rundle, "Use of Google Street View to assess environmental contributions to pedestrian injury," *American journal of public health*, vol. 106, no. 3, pp. 462–469, 2016.

[25] C. E. Kontokosta, N. Johnson, and A. Schloss, "The quantified community at Red Hook: Urban sensing and citizen science in low-income neighborhoods," *arXiv preprint arXiv:1609.08780*, 2016.

[26] C. E. Kontokosta, "The quantified community and neighborhood labs: A framework for computational urban science and civic technology innovation," *Journal of Urban Technology*, vol. 23, no. 4, pp. 67–84, 2016.

[27] E. L. Glaeser, S. D. Kominers, M. Luca, and N. Naik, "Big data and big cities: The promises and limitations of improved measures of urban life," *Economic Inquiry*, vol. 56, no. 1, pp. 114–137, 2018.

[28] Mastercard Advisors. (2018) Mastercard data & analytics. [Online]. Available: https://www.mastercardadvisors.com/en-us.html

[29] Y. Zheng, "Methodologies for cross-domain data fusion: An overview," *IEEE transactions on big data*, vol. 1, no. 1, pp. 16–34, 2015.

[30] Y. Zhu and J. Ferreira, "Data integration to create large-scale spatially detailed synthetic populations," in *Planning Support Systems and Smart Cities*. Springer, 2015, pp. 121–141.

[31] C. M. Croner, J. Sperling, and F. R. Broome, "Geographic information systems (GIS): new perspectives in understanding human health and environmental relationships," *Statistics in medicine*, vol. 15, no. 18, pp. 1961–1977, 1996.

[32] M. Tresidder, "Using GIS to measure connectivity: an exploration of issues," *School of Urban Studies and Planning, Portland State University: Portland, OR, USA*, 2005.

[33] R. C. Brownson, C. M. Hoehner, K. Day, A. Forsyth, and J. F. Sallis, "Measuring the built environment for physical activity: state of the science," *American journal of preventive medicine*, vol. 36, no. 4, pp. S99–S123, 2009.

[34] R. Schneider, L. Arnold, and D. Ragland, "Pilot model for estimating pedestrian intersection crossing volumes," *Transportation Research Record: Journal of the Transportation Research Board*, no. 2140, pp. 13–26, 2009.

[35] K. M. Neckerman, M. Purciel-Hill, J. W. Quinn, and A. Rundle, "Urban design qualities for New York City," in *Measuring Urban Design*. Springer, 2013, pp. 63–82.

[36] M. F. Goodchild, L. Anselin, and U. Deichmann, "A framework for the areal interpolation of socioeconomic data," *Environment and planning A*, vol. 25, no. 3, pp. 383–397, 1993.

[37] J. Maantay, "Asthma and air pollution in the bronx: methodological and data considerations in using GIS for environmental justice and health research," *Health & place*, vol. 13, no. 1, pp. 32–56, 2007.

[38] B. Hillier, A. Leaman, P. Stansall, and M. Bedford, "Space syntax," *Environment and Planning B: Planning and Design*, vol. 3, no. 2, pp. 147–185, 1976.

[39] C. Ratti, "Space syntax: some inconsistencies," *Environment and Planning B: Planning and Design*, vol. 31, no. 4, pp. 487–499, 2004.

[40] L. D. Frank, T. L. Schmid, J. F. Sallis, J. Chapman, and B. E. Saelens, "Linking objectively measured physical activity with objectively measured urban form," *American journal of preventive medicine*, vol. 28, no. 2, pp. 117–125, 2005.

[41] NYC Department of Environmental Protection. (2018) Air pollution. [Online]. Available: https://www1.nyc.gov/site/dep/environment/air-pollution-regulations.page

[42] D. Tasse and J. I. Hong, "Using user-generated content to understand cities," in *Seeing Cities Through Big Data*. Springer, 2017, pp. 49–64.

[43] S. Doheny-Farina, *The wired neighborhood*. Yale University Press, 1998.

[44] G. District, "Pedestrian counts, the Garment District NYC," The Garment District Alliance, 2016. [Online]. Available: http://garmentdistrict.nyc/real-estate/pedestrian-counts/

[45] TimeSquare Alliance. (2016) Time Square Alliance market research data. Time Square Alliance. New York City. [Online]. Available: http://www.timessquarenyc.org

[46] NYC Department of Transportation, "Bi-annual pedestrian count data feed," New York City, 2016. [Online]. Available: http://www.nyc.gov/html/dot

[47] Department of Transportation. (2015) Vision zero: Pedestrian safety plan. [Online]. Available: http://www.nyc.gov

[48] Grand Central Partnership. (2018) Report and data: Pedestrian count. Grand Central Partnership. New York City. [Online]. Available: http://www.grandcentralpartnership.nyc/about/reports-and-data