Original Research Article

Decoding the association between urban streetscape skeletons and urban activities: Experiments in Beijing using Dazhong Dianping data

Enjia Zhang
School of Architecture, Tsinghua University, China

Hanting Xie
School of Architecture, Tsinghua University, China
AECOM San Francisco Office, USA

Ying Long
School of Architecture, Tsinghua University, China
Key Laboratory of Eco Planning & Green Building, Ministry of Education, China

Abstract

The essential role of urban streetscape skeletons in fostering vibrant streets has been repeatedly emphasized. However, most research focused on the macro-scale or mesoscale urban form or measuring streetscape skeletons using quantitative methods, failing to systematically examine the relationship between the street-level urban form with urban activities. This study took street segments as the analytic unit to analyze the relationship between streetscape skeletons and urban activities represented by the density of online reviews on the Dazhong Dianping by controlling other built environments. Using models like Multiple Linear Regression, Spatial Lag Model, and Random Forest Regression, the result suggested that when conducting activities, people preferred the strengths of the street but ignored its weaknesses. The findings demonstrated some skeleton indicators associated with urban activities, such as the width and enclosure of the street and the higher and more continuous buildings on the side of the streets. Moreover, the result also suggested that some streetscape skeletons, such as cross-section and length, have differentiated performances for online reviews at different distances to the street centerlines. These findings could help urban designers to rethink the interaction between urban activities and street-level urban form.

Keywords

Social media data, streetscape skeleton, street vitality, urban activity, urban form

Corresponding author:
Ying Long, School of Architecture, Tsinghua University, 1 Qinghuayuan Sub-district, Haidian District, Beijing 100084, China.
Email: ylong@tsinghua.edu.cn

Introduction

The essential role of urban streetscapes in fostering vibrant streets has been repeatedly emphasized in conventional urban design guidelines (Tucker et al., 2005). The streetscape skeleton, as the three-dimensional spaces of streets defined by the size and arrangement of large neighboring objects such as buildings and trees, is an irreplaceable part of street design to create prolific public spaces and accelerate abundant daily activity (Harvey and Aultman-Hall, 2016). Livable streetscape skeletons can help citizens construct impressive city images (Lynch, 1960) and instill a sense of place by providing attractive and comfortable outdoor spaces for staying and communicating (Gehl, 2011). Therefore, there have been numerous attempts to explore suitable streetscape skeletons for urban designers to better understand human-scale urban form and shape vibrant and sticky streets in the past six decades (Ashihara, 1983; Jacobs, 1961; Long and Ye, 2019; Sternberg, 2000).

Recent advancements in the new data environment, such as street view pictures (SVP), point of interest (POI), open street maps (OSM), and social media data, have created a new context for gaining insights into human-scale streetscapes (Harvey et al., 2017; Harvey and Aultman-Hall, 2016; Long and Ye, 2019; Lu, 2018). A growing number of scholars are being committed to proposing various approaches to measuring streetscapes such as using GIS (Geographic Information System)-based methods (Harvey, 2014; Harvey and Aultman-Hall, 2016; Harvey et al., 2017) to depict the skeletons of the streetscape, and applying deep learning algorithms to extract visual elements from SVP data (Zhang et al., 2018). Meanwhile, some studies have also revealed the close relationship between streetscapes and urban activities, such as diverse functions (Yoshimura et al., 2021), short streets (Long and Huang, 2019; Sung et al., 2015), and street-level greenery (Lu, 2018). However, the association between urban activities and geometric arrangements in the streets that urban designers also consider when discussing streetscapes, such as cross-sectional proportion, the continuity of urban streets (Ashihara, 1983), enclosure and spaciousness (Stamps, 2009), and the average height of buildings (Harvey, 2014), have not been systematically analyzed and discussed.

Are the streetscape skeletons frequently highlighted in conventional urban design guidelines associated with urban activities? What three-dimensional street forms are pedestrians prefer when they engage in urban activities? To address these two questions, we took social media data—online reviews—from the Dazhong Dianping platform (China's version of Yelp) as a proxy for people's in-person activities and willingness to share or review some locations or places. The relationship between the density of online reviews and streetscape skeletons was examined in Multiple Linear Regression (MLR) by controlling other indicators that could influence travel demand. The Spatial Lag Model (SLM) and Random Forest Regression (RFR) were used to verify the robustness of the results.

Literature review

Built environment potential to moderate travel demand

The built environment that could influence travel demand is a hot research topic in urban planning and transportation. Early explorations based on subjective perception and evaluation provided a fundamental framework for a sustainable and preferred built environment via interviews, in-person questionnaires, and field observations (Ashihara, 1983; Jacobs, 1961, 1995; Lynch, 1960). These subjective perceptions are the subjective impression or image of the human-scale urban form and built environmental elements for citizens, auditors, or spectators, such as aesthetics (Ashihara, 1983), safety, vitality (Jacobs, 1961), imageability (Lynch, 1960), enclosure, comfort (Gehl, 2011), transparency, and complexity (Ewing and Handy, 2009). Some scholars summarized the influential built environment features with words beginning with D through literature reviews to describe pedestrian activity-related built environments. Initially, "three Ds" (density, diversity, and design) were proposed (Cervero and Kockelman, 1997) to emphasize the essential role of urban form, land use, and design features in appealing travel activities. Afterward, two Ds- destination

accessibility and distance to transit- were introduced to underline the transport accessibility defined by the transport links with other vital nodes or landmarks (Ewing and Cervero, 2010).

Recently, the emergence of social sensing data (Liu et al., 2015) and open digital map data has enabled researchers to "measure the unmeasurable" (Ewing and Handy, 2009) and verify the significance of five Ds by investigating their relationship with urban activities (Long and Huang, 2019; Meng and Xing, 2019; Xia et al., 2020). Some indicators such as transport and landmark accessibility (Sung et al., 2015), mixed land use (Frank et al., 2006; Grant, 2002), intersection density (Ewing and Cervero, 2010; Sung et al., 2015), and street-level greenery (Lu et al., 2018) were verified highly associated with urban activities. Among various elements in five Ds, the design features vary the most among different spatial scales and analytic units. Meanwhile, the street-level design elements, coined "streetscapes", are fundamental objects that urban designers and architects emphasize in urban design guidelines. Therefore, compared with researchers in the transportation field who focus more on transportation networks and land use, urban planning and design scholars also highlight human-scale design features. However, most studies examined the urban form at macro and meso scales by using blocks or grids as analytic units, failing to ascertain the relationship between urban activities and urban design elements at the street level.

Streetscapes features related to urban activities

The emphasis on the street system can be traced back to Lynch's city images, where paths are regarded as channel-like images that contain physical movement and contribute to the urban configuration (Lynch, 1960). Then, Jacobs (1961) proposed four conditions of urban design and underlined the critical significance of livable streets and small blocks. Later, the reciprocal relationships between specific urban forms and activities were described in Gehl's book (Gehl, 2011), in which Gehl examined the relationship between patterns of space use and spatial properties of the physical environment. Some more specific indicators, such as the relationship between buildings and streets and the aspect ratio, were proposed by Ashihara (1983) to discuss the geometric structure of streets. With the emergence of space syntax, numerous researchers started to regard streets as linked linear spatial units that contribute to the functional organization of cities (Hillier, 2012; Shen and Karimi, 2016).

In this context, streetscapes were introduced to describe the human-scale built environmental characteristics of streets (Tang and Long, 2019), which are defined as the spatial arrangement, street furniture, and visual appearance of built and landscape features when viewed from the street (Tucker et al., 2005). Ewing et al. (2006) identified and measured urban design characteristics via an audit protocol, which inspired Purciel et al. (2009) to measure 22 physical features in the streets in terms of imageability, enclosure, human scale, transparency, and complexity using GIS-based methods. Referring to Purciel's proposal, Ewing and Handy (2009) summarized 20 streetscapes from video clips regarding five design qualities. Rehan (2013) introduced 17 streetscape elements to describe the natural and built fabric of the street, ranging from buildings and street surfaces to fixtures and fittings that facilitate its use. Harvey (2014) used the term "streetscape skeleton" to depict the 12 characteristics of spatial arrangement in three-dimensional space to show the size and arrangement of buildings and trees within streetscapes, and in 2016, Harvey and Aultman-Hall (2016) proposed a framework to illustrate 19 components of streetscapes. Although these scholars focused on different dimensions of streetscapes from diversified perspectives, they all emphasized the importance of streetscape skeletons, street furniture, physical and visual appearance, and land use.

Recently, emerging big data such as open street maps and street view pictures, and deep learning algorithms have enabled an increasing number of studies on quantifying streetscapes. (Li, 2021; Ye et al., 2019). Some studies applied the GIS method to depict the three-dimensional geometric characteristics of the street and surrounding buildings and trees (Harvey et al., 2017; Purciel et al., 2009). Meanwhile, some studies explored extracting visual elements from street-view pictures. Among various deep learning techniques, image classification, object detection, and semantic segmentation were three frequently used models for
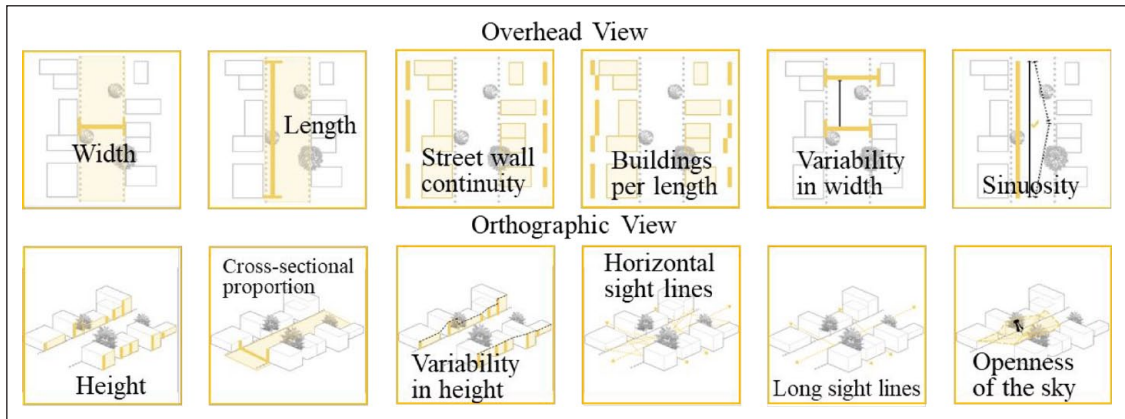
Figure 1. Streetscape skeletons define the three-dimensional space of a street.

extracting information from street view pictures, which has a broad range of potential applications (Ibrahim et al., 2019; Kruse et al., 2021; Tang and Long, 2019). Although several studies revealed some streetscapes highly associated with pedestrian activities, such as street-level greenery, openness, and spaciousness, they failed to systematically identify the associated streetscape skeletons.

Streetscape skeleton and its measurement

Academic discussions on the qualitative and quantitative measurement of streetscape skeletons have evolved over the past 60 years. Jacobs (1961) underlined the necessity of small blocks (short streets) based on subjective perception and observation. Ashihara (1983) introduced a quantified index - the width-to-height ratio (D/H) - to define the aesthetics of the street by measuring the relationship between buildings and streets. Measuring skeleton indicators using GIS-based methods is a state-of-the-art method that examines three-dimensional geometric relationships based on vector data with geographic information. The street wall continuity, the average height of buildings, and the count of buildings along the street were measured in most studies (Ewing and Handy, 2009; Harvey, 2014; Purciel et al., 2009). Purciel et al. (2009) and Ewing et al. (2015) also emphasized the sight range in the street, simulated the horizontal sight lines, and counted the number of long sight lines and the proportion of the visible sky, respectively. In comparison, Harvey (2014) introduced more specific measures of street skeletons, such as width between buildings across the street, cross-sectional proportion, length of centerline, variability in height and width, and sinuosity of the centerline, to portray the arrangement of buildings along either side of a streetscape (Figure 1). However, most studies on streetscapes focused on the measurement rather than their relationship with urban activities. Therefore, there is an urgent need to understand the streetscapes repeatedly cited in previous studies and recognize elements that urban designers need to focus on during urban design.

Methodology

Research framework

This study focused on Beijing's Fifth Ring Road area, a central city area covering approximately $667\,km^2$. To achieve a deeper understanding of the relationship between streetscape skeletons and urban activities, we used streets as the analytic unit to measure the human-scale urban form. The raw data for this investigation contained 17,330 streets with street-level spatial data within the Fifth Ring Road area (Figure 2).
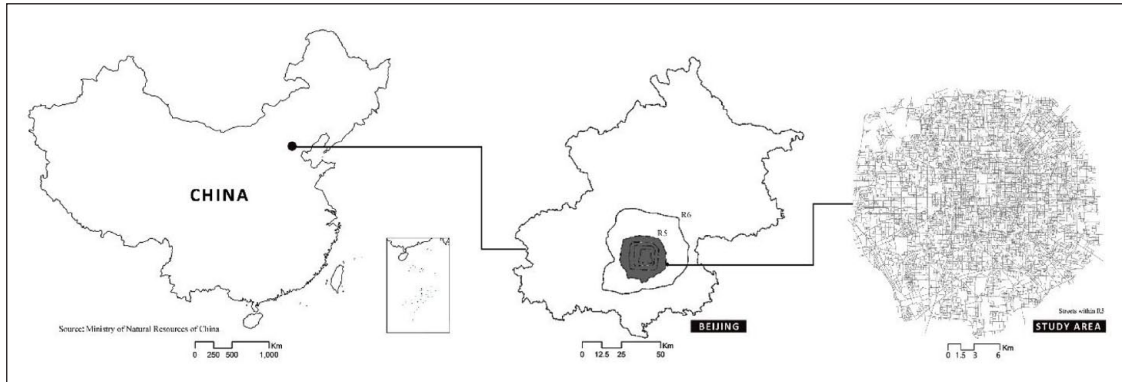
Figure 2. Study area.

In this research, we took online reviews on Dazhong Dianping (the most prominent online rating and deal service platform for local recreation providers in China) (Long and Huang, 2019) as a proxy of urban activities for three reasons (see Supplementary Information (SI) for Figure S1). First, online reviews were highly associated with on-site consumption, and thus could reflect the actual in-person activities. Second, online reviews reflected the consumer's preference to share their opinions. Though there might be some bad reviews or provider-induced good reviews, only service providers with good quality could receive more online reviews. Therefore, the number of good reviews also closely correlated with the total number of online reviews. Third, the online reviews could serve as the Electronic Word of Mouth (eWoM) (Cheung and Lee, 2012), which could influence customers' demand confirmation, location choice, and post-evaluation (Mu et al., 2019; Shi et al., 2021) thus has the potential to attract more pedestrian flows (Xu et al., 2019).

We used the density of online reviews as the dependent variable to examine its association with 11 streetscape skeletons (width, length, height on the higher and lower side, cross-section on the higher and lower side, continuity of the higher and lower side, buildings per length, sinuosity, and openness of the sky) by controlling for the other four Ds—destination accessibility (distance to city center), distance to transit (distance to metro station), density (density of various POIs), and diversity (mixing degree of POIs). Multiple Linear Regression (MLR)—Ordinary Least Square (OLS) Estimation—was employed in this study to conduct step-wise regressions after examining the difference in results with different radii of the buffer for the street to calculate the density of online reviews. The Spatial Lag Model and Random Forest Regression were applied to test the robustness of the result (Figure 3). Based on the standard linear regression, SLM considered the autocorrelation among observed variables by generating the matrix between streets (Seya et al., 2020). We generated a 50-m buffer of streets and ran the SLM in GeoDa. The RFR was also used because it could provide an index—the proportion of variance explained (Prop. Var)—for the overall model and a series of "variable importance" scores for predictor variables (Smith et al., 2013). The Cross Validation was used to automatically and randomly take a proportion of different data as test and train data to obtain the average Prop. Var.

Measurement of the dependent variable—urban activity

The density of online reviews derived from Dazhong Dianping was considered the dependent variable representing urban activity. Online reviews were generated from consumers who scored and commented on places or service providers they patronized. From the company's website (http://www.dianping.com), we obtained available information on service providers, including the name, detailed location (longitude and latitude), and the number of reviews (see Figure S2 in SI for details). After data deduplication, we converted
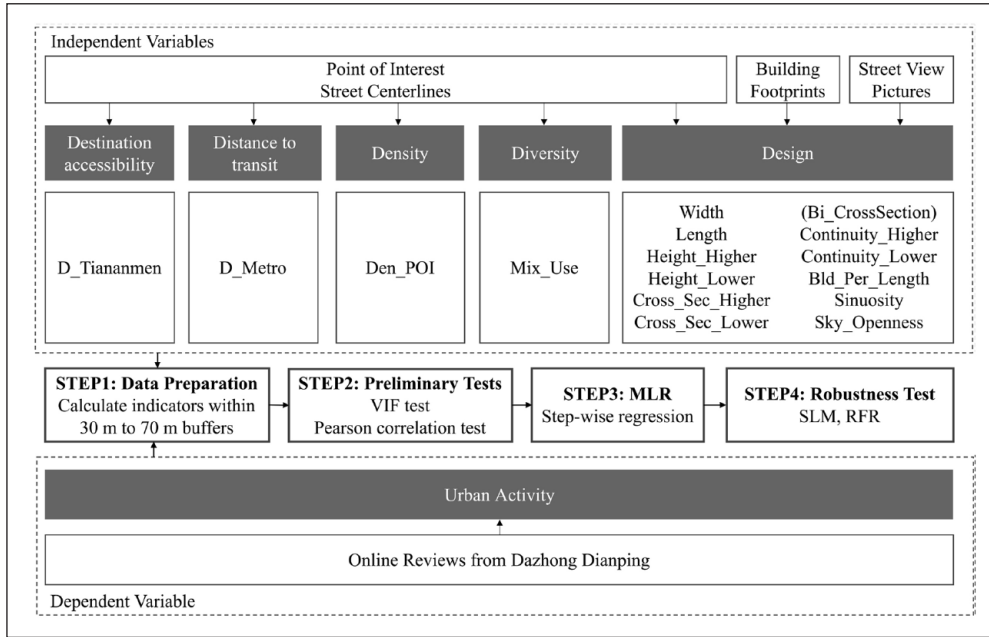
Figure 3. Analytical framework.

the address to the WGS-84 coordinate to ensure all the data was in the same coordinate. Then, to avoid the impact of the buffer area and the length of the street on the results, we calculated the density of online reviews within a buffer zone for each street as the value of each sample. Given that the radius to generate the buffer of the street might interfere with the results, this study selected the radius 30 m, 40 m, 50 m, 60 m, and 70 m to build the buffer, respectively, and compared their differences. According to the spatial visualization of the density of Dianping POIs and online reviews in the 50 m buffer of the street, the distribution of Dianping POIs was even across the city, while the density of online reviews clustered around the north-eastern part of the city (Figure S3 in SI).

Measurement of the independent variable—streetscape skeletons

The streetscape skeletons were regarded as the design indicators in this study. We acquired building footprint data with layer information and street centerlines to calculate skeleton indicators according to methods developed by Harvey (2014). We also applied a SegNet method to count the visual proportion of the sky in street-view pictures (Note 1 in SI) to depict the openness of the sky, referring to Ewing and Handy (2009). All the streetscape skeletons included the width (Width), length (Length) and sinuosity (Sinuosity) of the street, and the height (Height_Higher and Height_Lower), cross-section (Cross_Sec_Higher and Corss_Sec_Lower), continuity (Continuity_More and Continuity_Less), and density (Bld_Per_Length) of buildings on two sides, and the openness of sky from the view of the street (Sky_Openness). Considering the existing ideal ranges for cross-sectional proportions (width/height) in the conventional urban design theory, we further transferred the cross-section variables into a binary variable in which the proportion values between the ideal ranges (1 and 2 according to Ashihara (1983)) were transferred into 1, and other proportion values were transferred into 0. The detailed descriptions and methods to calculate these skeleton indicators are presented in Table 1.

Table 1. Dependent and independent variables and their descriptions.

Indicator system	Variable names	Descriptions	Data sources	Mean	Std_deviation	Unit
Urban activities	R30mDen_Review	The density of online reviews within each street's 30m to 70m buffer.	Dazhong Dianping	3.12	1.87	#/km²
	R40mDen_Review			2.80	1.96	#/km²
	R50mDen_Review			3.33	1.78	#/km²
	R60mDen_Review			3.48	1.70	#/km²
	R70mDen_Review			3.58	1.63	#/km²
Design	Width	Distance between edges (building-to-building) across the street.	Amap	35.30	18.71	m
	Length	Centerline distance between intersections.	Amap	248.95	181.20	m
	Height_Higher	Average building height on the higher or lower side of the street.	Amap	16.64	17.25	m
	Height_Lower		Amap	6.84	8.96	m
	Cross_Sec_Higher	Full width (building-to-building) / height on the higher or lower side of the street.	Amap	0.53	0.76	—
	Cross_Sec_Lower		Amap	0.22	0.45	—
	Bi_CrossSection	The average street-level cross-sectional proportion of each street. The proportion values ranging between 1 and 2 were transferred into 1; otherwise, the values were transferred into 0.	Amap	0.17	0.38	—
	Continuity_Higher	The proportion of edge intersecting buildings on the more or less continuous side of the street.	Amap	0.44	0.27	—
	Continuity_Lower		Amap	0.23	0.23	—
	Bld_Per_Length	Count of buildings on both sides / street length.	Amap	23.10	16.98	#/km
	Sinuosity	Centerline length / straight line distance between segment ends.	Amap	1.08	0.39	—
	Sky_Openness	The street-level proportion of the sky was calculated by averaging the SegNet proportions in all the street view pictures on the corresponding street.	Tencent Map	0.16	0.09	—
Destination accessibility	D_Tiananmen	The Euclidean distance from each street midpoint to Tiananmen (The flagpole point on Tiananmen Square).	Amap	8.66	3.62	km
Distance to transit	D_Metro	The road network distance from the street midpoint to the closest subway station (the 222 subway stations within our study area).	Amap	1.68	1.30	km
Density	Den_POI	The density of POIs within the 50-m buffer of each street.	Amap	2674.16	4898.92	#/km²
Density	Mix_Use	The mixing degree of POIs was calculated using the formula $mix\_use = -sum(p_i^{\ast}ln\ p_i)$, ($i=1,2,..,n$) where n represents the number of POI types on the street and pi represents the standardized proportion of a specific POI type on the street.	Amap	1.44	0.63	—
Location	Within_R2	Dummy variables for the location of each street. If the street is in the subregion, the value is 1; otherwise, the value is 0.	Amap	—	—	—
	R2_R3			—	—	—
	R3_R4			—	—	—
	R4_R5			—	—	—

Road networks, building footprints, and POI (point of interest) data were collected from Amap (one of the most popular and reliable driving navigation apps, https://lbs.amap.com) from its open and official API (application program interface).

Measurement of the independent variable—control variables

Other four types of elements in the five Ds were used as the control variables in this study. This study applied four elements that have been verified highly associated with urban activities. (1) Destination accessibility. The distance to the central city was the accessibility indicator to describe the specific location of each street by calculating the Euclidean distances between the midpoint of the street to the flagpole point based on Tiananmen Square. (2) Distance to Transit. The distance to transit was measured by the network distance from the street midpoint to the nearest metro station. (3) Density. The density of various functions was selected as the density indicator, which was the POI (point of interest) density calculated within the defined radius of the buffer of each street. (4) Diversity. As for the diversity indicator, this study examined the mixing degree of POIs within the buffer zone for each street using Shannon's entropy. Specifically, the POIs were firstly divided into six types of land use, namely, residential, commercial, industrial, public management and public service, road, and transportation facilities, and park and green land. Then, the normalized proportion of each type of POIs was calculated with Shannon's entropy. The final result represented the degree of mixed land use, where a value close to 1 showed the street had more diversified uses. Besides, regarding the different functions of the different ring roads of Beijing, this study added the dummy variable to indicate the subregions of each street: 1) within the 2nd Ring Road areas, 2) between the 2nd and 3rd Ring Road areas, 3) between the 3rd and 4th Ring Road areas; and 4) between the 4th and 5th Ring Road areas. Table 1 shows the descriptions and methods to calculate the aforementioned control variables. The spatial visualization of independent variables is shown in Figure S4 in SI.

Results

Data processing and preliminary tests

Because the OLS regression assumes that the involved variables are normally distributed, we examined the dependent variables' original data distribution through histograms and boxplots. The variables of online review density and POI density were long-tailed. Therefore, we used the log([VAR]+1) transformation on the density of online reviews and removed zero values in the subsequent regression to guarantee the model's reliability. Besides, we also preprocessed the independent variables and applied logarithmic transformation for the variable Den_POI. After logarithmic transformation, we examined the data distribution to check whether this step would help achieve normal distribution. After data cleaning, the final sample size was 16,718.

Before completing the regressions, we employed the Pearson correlation and VIF (variance inflation factor) tests to determine the regression models' reliability and avoid the multicollinearity effect. Since the dummy variable R3_R4 was highly correlated with other location variables, this study removed this indicator. Then, the results showed that the correlation coefficients between other variables were less than 0.8, and the VIF values were less than 5. That means the multicollinearity of the model was not very severe, and the model was relatively reliable.

Multiple Linear Regression analysis and results

Before we compared the differences between streetscape skeletons and other indicators, we examined the influence of the selected radius to generate the buffer of the street and to calculate the density of online reviews on the robustness of the results. The results showed that (Table 2), with the buffer radius raised from 30 m to 50 m to measure the aforementioned indicators, the number of elements significantly associated with online reviews decreased; when the radius was higher than 50 m, results tended to present similar patterns. According to the results, most skeleton indicators were significantly associated with urban activities using 30 m and 40 m as the radius of the buffer. One possible implication of this result is that urban activities

Table 2. MLR results using different radii to generate buffers and calculate indicators.

N = 16,718	30 m buffer	40 m buffer	50 m buffer	60 m buffer	70 m buffer	VIF
Within_R2	**−0.0289****	−0.0185	**−0.0378*****	**−0.0438*****	**−0.0482*****	2.8930
R2_R3	**−0.0549*****	**−0.0492*****	**−0.0564*****	**−0.0557*****	**−0.0533*****	1.9348
R4_R5	0.0163	0.0005	0.0161	0.0168	0.0083	2.7368
D_Tiananmen	**−0.1731*****	**−0.1409*****	**−0.1957*****	**−0.2192*****	**−0.2305*****	5.4625
D_Metro	**−0.0366*****	**−0.031*****	**−0.0423*****	**−0.0506*****	**−0.0547*****	1.1784
LnDen_POI	**0.2888*****	**0.2716*****	**0.3091*****	**0.3245*****	**0.3334*****	2.3305
Mix_Use	**0.2474*****	**0.232*****	**0.2492*****	**0.2293*****	**0.21*****	2.2778
Width	**0.0591*****	−0.0018	**0.086*****	**0.0913*****	**0.0962*****	1.4811
Length	**0.0139***	**0.0428*****	−0.0069	**−0.0245*****	**−0.0336*****	1.2710
Height_Higher	**0.0493*****	**0.0316****	**0.0708*****	**0.0785*****	**0.0813*****	3.7696
Height_Lower	**0.0202***	**0.0322****	−0.0002	−0.0083	−0.0094	3.1580
Cross_Sec_Higher	**0.0345****	**0.0356****	0.0197	0.0171	**0.0215***	4.0832
Cross_Sec_Lower	**−0.0225***	**−0.0278****	−0.0091	−0.0032	−0.0095	3.4354
Continuity_Higher	**0.092*****	**0.0878*****	**0.0699*****	**0.0504*****	**0.0402*****	2.5178
Continuity_Lower	**0.0363*****	**0.0748*****	0.0122	0.0041	−0.007	2.6225
Bld_Per_Length	**−0.0162***	−0.0013	**−0.0305*****	**−0.0456*****	**−0.054*****	2.1070
Sinuosity	−0.0042	−0.0069	−0.0023	−0.0022	−0.0007	1.0754
Sky_Openness	**−0.0971*****	**−0.0919*****	**−0.1069*****	**−0.1111*****	**−0.1192*****	1.7024
R^2	0.5108	0.4613	0.5364	0.5356	0.5356	
Adjusted R^2	0.5103	0.4607	0.5359	0.5351	0.5351	

Each column was a different model. The table reports the model's predictive power (R^2) for each model's column. The coefficients in the regression results were standardized, within which the bolded variables were statistically significant (<0.05). For statistical significance, we used the following notation: *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

represented by the density of online reviews near the centerlines were highly associated with more skeleton variables, such as those in living and commercial walking streets. When some service providers further away from the centerlines were analyzed in larger buffers, skeleton variables, such as length and cross-section indicators, had adverse results or were not statistically significant. Since the 50-m buffer had similar results to buffers with higher radii, this study chose the 50-m buffer to conduct the following studies and robustness tests.

Then, we conducted a step-wise OLS regression to explore the explanatory power of different Ds (Table 3). The results showed less density of urban activities within the third ring road of Beijing and verified previous findings that destination accessibility, distance to transit, density, and diversity had close relationships with urban activities. Among these, the density and diversity indicators had higher positive and statistically significant coefficients. As for the streetscape skeletons - the design indicators in this study - the width of the street, the higher height and continuity of buildings along the street were positively and significantly associated with online reviews, indicating that the broad street with high-rise buildings and more continuous street walls are more likely to witness more activities. Meanwhile, insignificant coefficients of Height_Lower and Continuity_Lower suggested that urban activities preferred the strength of the street, such as the higher height and continuity, but ignored the drawbacks of the other side. Besides, this study revealed the negative correlations between building density and the openness of the sky in the street with urban activities. In terms of the urban form and layout of buildings in Beijing, streets with more buildings along the street are north-south streets, which are along the side face of the buildings, thus with fewer storefronts or street-level ground-floor windows. The higher proportion of visual sky in the street view picture indicated fewer trees and buildings in the street, which might not be suitable enough for people to conduct activities. Moreover, although this study also used the dummy variable to represent the ideal cross-section as an alternative, the cross-section indicators that conventional urban design highlighted were not statistically associated with urban activities.

Table 3. The results of step-wise MLR for different dimensions of variables.

N = 16,718	Variable description	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Location	Within_R2	**−0.1207*****	**−0.103*****	**−0.065*****	**−0.0605*****	**−0.0378*****	**−0.0378*****
	R2_R3	**−0.0715*****	**−0.0576*****	**−0.0444*****	**−0.0497*****	**−0.0564*****	**−0.0562*****
	R4_R5	**−0.0523*****	**−0.0423*****	**0.0301****	**0.0211***	0.0161	0.0163
Destination accessibility	D_Tiananmen	**−0.4543*****	**−0.4049*****	**−0.2467*****	**−0.2258*****	**−0.1957*****	**−0.1956*****
Distance to transit	D_Metro		**−0.1195*****	**−0.0561*****	**−0.0546*****	**−0.0423*****	**−0.0422*****
Density	LnDen_POI			**0.5696*****	**0.3573*****	**0.3091*****	**0.3092*****
Diversity	Mix_Use				**0.3176*****	**0.2492*****	**0.2495*****
Design (Streetscape skeleton)	Width					**0.086*****	**0.0831*****
	Length					−0.0069	−0.0067
	Height_Higher					**0.0708*****	**0.0808*****
	Height_Lower					−0.0002	−0.0061
	Cross_Sec_Higher					0.0197	
	Cross_Sec_Lower					−0.0091	
	Bi_CrossSection						0.0098
	Continuity_Higher					**0.0699*****	**0.0705*****
	Continuity_Lower					0.0122	0.0111
	Bld_Per_Length					**−0.0305*****	**−0.0306*****
	Sinuosity					−0.0023	−0.002
	Sky_Openness					**−0.1069*****	**−0.1063*****
	R^2	0.1614	0.1738	0.4479	0.5013	0.5364	0.5364
	Adjusted R^2	0.1612	0.1735	0.4477	0.5011	0.5359	0.5359

Each column was a different model. The table reports the model's predictive power (R^2) for each model's column. The coefficients in the regression results were standardized, within which the bolded variables were statistically significant ($<$0.05). For statistical significance, we used the following notation: *p $<$ 0.05, **p $<$ 0.01, ***p $<$ 0.001.

Overall, the result using the density of online reviews on Dazhong Dianping as the dependent variable uncovered the close relationship between some skeleton indicators, such as width, higher height, and more continuous side of the street with urban activities. Nevertheless, some indicators such as Cross_SecHigher, Corss_SecLower, and Length have different performances when considering different buffer areas. Besides, this study also indicated that streetscape skeletons have lower explanatory power than function-related indicators, such as the density and diversity of POIs.

Robustness tests

Results using Spatial Lag Model

Before using the spatial lag model, we conducted a global Moran's I test for the dependent variable to examine the degree of spatial autocorrelation. The result showed that the p-values were significantly less than 0.05, the z-values were significantly larger than 1.96, and Moran's I was positive, indicating that the online review density showed significantly clustered patterns. Therefore, we applied the SLM to better understand the clusters of urban activities. The results showed that MLR and SLM could explain 53.62% and 53.82% of online reviews, respectively, suggesting that SLM had a slightly better performance than MLR, and implying the spatial autocorrelation of the density of online reviews was apparent. The results also showed that by controlling the spatial autocorrelation effect, all the dependent variables showed similar performances to MLR, indicating that the results in the above study were reliable (Table 4).

Results using Random Forest Regression

This study also used 10-fold Cross-Validation in RFR to test the reliability of the results from MLR. All the observations were automatically divided into training and test sets (90%:10%) to run the model and test its accuracy. This process was repeated ten times using different training and test data, resulting in the average value of the proportion of variance explained (Prop. Var).

The Adjusted R^2 values for the MLR were compared to the Prop. Var values for the RFR (Table 5). The proportion of Prop. Var in RFR ranged from 0.49 to 0.55, and the Adjusted R^2 values in the MLR were slightly lower, varying from 0.46 to 0.54. Figure 4 shows the importance of variables. For the RFR, these variables were the ones that had the largest effect on the MSE (mean square error), and for the MLR, they were the statistically significant variables, listed in order from the largest to the smallest absolute value of the standardized coefficients. The same number of variables is shown for the RFR as for the MLR: the number of variables with significant coefficients (p-value <0.05) in MLR. In the two regressions, function density, diversity, and distance to the city center were the most significant and important variables. As for skeleton variables, although the orders varied, the significance of the higher side of the street, openness of the sky, street width, building density, and higher continuity of buildings along the streets was revealed in two regressions. Overall, the similar Adjusted R^2 in MLR and Prop. Var in RFR, and similar variables with higher importance indicated that the results in this study were reliable.

Discussion

Using the street segment as the analytic unit, this study focused on the human-scale urban design elements—streetscape skeletons, which may provide more insights into the refined management of the built environment. Some conclusions of this research could enhance the understanding of diverse urban design ideas and reinforce urban design pedagogy and the theoretical foundation of urban design.

On the one hand, the results verified previous scholars' claims by revealing the importance of some streetscape skeletons. The results suggested that when conducting activities, people preferred the strengths

**Table 4.** The comparison between MLR and SLM.

N = 16,718	MLR	SLM
W_R50mDen_Review		**0.0410*****
Within_R2	**−0.1891*****	**−0.1766*****
R2_R3	**−0.2487*****	**−0.2413*****
R4_R5	0.0574	0.0496
D_Tiananmen	**−0.0001*****	**−0.0001*****
D_Metro	**−0.0001*****	**−0.0001*****
LnDen_POI	**0.7654*****	**0.7664*****
Mix_Use	**0.7025*****	**0.7006*****
Width	**0.0082*****	**0.0082*****
Length	−0.0001	−0.0001
Height_Higher	**0.0074*****	**0.0074*****
Height_Lower	−0.0002	−0.0001
Cross_Sec_Higher	0.0420	0.0417
Cross_Sec_Lower	−0.0306	−0.0306
Continuity_Higher	**0.4566*****	**0.4534*****
Continuity_Lower	0.0976	0.0968
Bld_Per_Length	**−3.2551*****	**−3.2761*****
Sinuosity	−0.0033	−0.0027
Sky_Openness	**−2.0290*****	**−2.0034*****
R^2	0.5362	0.5372

Each column was a different model. The table reports the model's predictive power (R^2) for each model's column. The bolded coefficients in the regression were statistically significant (<0.05). For statistical significance, we used the following notation: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Table 5. The comparison between adjusted R^2 in MLR and prop. var in RFR.

	30 m buffer	40 m buffer	50 m buffer	60 m buffer	70 m buffer
MLR-Adjusted R^2	0.51	0.46	0.54	0.54	0.54
RFR-Prop. Var	0.53	0.49	0.55	0.55	0.55

of the street but ignored its weaknesses. The street's higher and more continuous sides of buildings were associated with urban activities, while the lower and less continuous sides were not statistically significant. Besides, the positive coefficient of width and the negative one for the proportion of sky in the street implied that broad roads with higher enclosure are more likely to attract more urban activities in Beijing. On the other hand, the findings also provided some alternative perspectives on the streetscape skeletons. For example, the cross-section indicators were significantly associated with urban activities in narrow buffers of streets but were insignificant for wider ones. The relationship between street length and urban activities changed with more online reviews being involved in buffers to calculate the activity density. For those close to the centerline of streets, long streets received more activities, while for those further away from the centerline, short streets could be more appealing to pedestrians. These findings could enhance understanding of the complexity of the relationship between the three-dimensional street form (streetscape skeletons) and urban activities. Meanwhile, this study also compared the skeleton indicators with other elements. Besides, the results suggested that function density and diversity indicators played the most significant role in
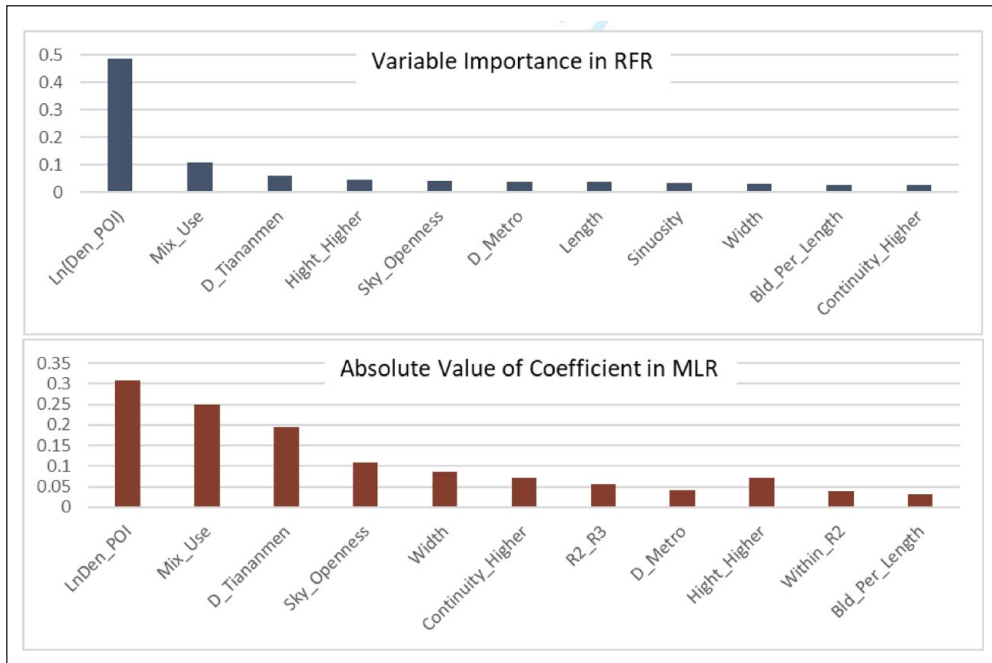
Figure 4. The variables in order of importance in the RFR and the absolute value of the coefficients in the MLR.

explaining the relationship between built environmental elements and urban activities. Urban designers should pay more attention to the guidance of diverse functions in urban renewal.

The present empirical analysis has several limitations. The first is the representativeness of the data. Online reviews represent only one specific dimension of urban activities and cannot wholly reflect people's preferences for built environment characteristics. Further studies are required to benchmark our analysis with other proxies, such as 5G location-based service (LBS) data for urban activity. Second, the analysis lacks a temporal dimension. While urban activity is time-dependent and varies in different months, weeks, days, and times of the day, the current time granularity of built environment data is insufficient to support the research on these different time sections. Information about urban activities and their performances throughout different times should be included in further studies to increase the robustness of the analyses. Third, some insignificant results and the generalizability of the significant results require more samples and variables for falsification or verification.

Conclusions

Human-scale urban form, as a basis for urban planning and design, can enhance the character of the built environment (Ewing et al., 2015; Long and Ye, 2019) and has the potential to attract more urban activities. Taking the central Beijing, China, as the study area, this study used the street segment as the analytic unit to examine the relationship between the street-level urban form-streetscape skeletons and urban activities represented by the density of online reviews from Dazhong Dianping through models like Multiple Linear Regression, Spatial Lag Model, and Random Forest Regression. The results verified some streetscape skeletons related to urban activities, such as the width and enclosure of the street and the higher and more continuous buildings on the side of the streets. Meanwhile, the findings also suggested that some skeleton indicators, such as cross-section and length, have differentiated performances in narrow and wide streets.

Besides, the results indicated that although some streetscape skeletons were highly associated with urban activities, function density and diversity were much more important. The findings of this study could provide new insights into streetscapes and help urban designers design more vibrant and sticky streets.

Supplemental material

Supplemental material for this article is available online.

References

Ashihara Y (1983) *The Aesthetic Townscape*. Cambridge, MA: The MIT Press.

Cervero R and Kockelman K (1997) Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment* 2(3): 199–219.

Cheung CMK and Lee MKO (2012) What drives consumers to spread electronic word of mouth in online consumer-opinion platforms. *Decision Support Systems* 53(1): 218–225.

Ewing R and Cervero R (2010) Travel and the built environment. *Journal of the American Planning Association* 76(3): 265–294.

Ewing R and Handy S (2009) Measuring the unmeasurable: Urban design qualities related to walkability. *Journal of Urban Design* 14(1): 65–84.

Ewing R, Handy S, Brownson RC, et al. (2006) Identifying and measuring urban design qualities related to walkability. *Journal of Physical Activity and Health* 3(s1): S223–S240.

Ewing R, Hajrasouliha A, Neckerman KM, et al. (2015) Streetscape features related to pedestrian activity. *Journal of Planning Education and Research* 36(1): 5–15.

Frank LD, Sallis JF, Conway TL, et al. (2006) Many pathways from land use to health: Associations between neighborhood walkability and active transportation, body mass index, and air quality. *Journal of the American Planning Association* 72(1): 75–87.

Gehl J (2011) *Life Between Buildings*. Washington, DC: Island Press.

Grant J (2002) Mixed use in theory and practice: Canadian experience with implementing a planning principle. *Journal of the American Planning Association* 68(1): 71–84.

Harvey CW (2014) *Measuring streetscape design for livability using spatial data and methods*. Graduate College Dissertations and Theses, University of Vermont, Vermont.

Harvey C and Aultman-Hall L (2016) Measuring urban streetscapes for livability: A review of approaches. *The Professional Geographer* 68(1): 149–158.

Harvey C, Aultman-Hall L, Troy A, et al. (2017) Streetscape skeleton measurement and classification. *Environment and Planning B-Urban Analytics and City Science* 44(4): 668–692.

Hillier B (2012) Studying cities to learn about minds: Some possible implications of space syntax for spatial cognition. *Environment and Planning B: Planning and Design* 39(1): 12–32.

Ibrahim MR, Haworth J and Cheng T (2019) URBAN-i: From urban scenes to mapping slums, transport modes, and pedestrians in cities using deep learning and computer vision. *Environment and Planning B: Urban Analytics and City Science* 48(1): 76–93.

Jacobs J (1961) *The Death and Life of Great American Cities*. New York, NY: Random House.

Jacobs AB (1995) *Great Streets*. Cambridge, MA: The MIT Press.

Kruse J, Kang Y, Liu Y-N, et al. (2021) Places for play: Understanding human perception of playability in cities using street view images and deep learning. *Computers, Environment and Urban Systems* 90: 101693.

Li X (2021) Examining the spatial distribution and temporal change of the green view index in New York City using google street view images and deep learning. *Environment and Planning B: Urban Analytics and City Science* 48(7): 2039–2054.

Liu Y, Liu X, Gao S, et al. (2015) Social sensing: A new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers* 105(3): 512–530.

Long Y and Huang CC (2019) Does block size matter? The impact of urban design on economic vitality for Chinese cities. *Environment and Planning B-Urban Analytics and City Science* 46(3): 406–422.

Long Y and Ye Y (2019) Measuring human-scale urban form and its performance. *Landscape and Urban Planning* 191: 103612.

Lu Y (2018) The association of urban greenness and walking behavior: Using google street view and deep learning techniques to estimate residents' exposure to urban greenness. *International Journal of Environment Research Public Health* 15(8): 1576.

Lu Y, Sarkar C and Xiao Y (2018) The effect of street-level greenery on walking behavior: Evidence from Hong Kong. *Social Science & Medicine* 208: 41–49.

Lynch K (1960) *The Image of the City*. Cambridge, MA: The MIT Press.

Meng Y and Xing H (2019) Exploring the relationship between landscape characteristics and urban vibrancy: A case study using morphology and review data. *Cities* 95: 102389.

Mu W, Spaargaren G and Oude Lansink A (2019) Mobile apps for green food practices and the role for consumers: A case study on dining out practices with Chinese and Dutch young consumers. *Sustainability* 11(5): 1275.

Purciel M, Neckerman KM, Lovasi GS, et al. (2009) Creating and validating GIS measures of urban design for health research. *Journal of Environmental Psychology* 29(4): 457–466.

Rehan RM (2013) Sustainable streetscape as an effective tool in sustainable urban design. *HBRC Journal* 9(2): 173–186.

Seya H, Yoshida T and Yamagata Y (2020) Chapter five-Spatial econometric models. In: Yamagata Y and Seya H (eds) *Spatial Analysis Using Big Data*. London: Academic Press, 113–158.

Shen Y and Karimi K (2016) Urban function connectivity: Characterisation of functional urban streets with social media check-in data. *Cities* 55: 9–21.

Shi Y, Tao T, Cao X, et al. (2021) The association between spatial attributes and neighborhood characteristics based on Meituan take-out data: Evidence from Shanghai business circles. *Journal of Retailing and Consumer Services* 58: 102302.

Smith PF, Ganesh S and Liu P (2013) A comparison of random forest regression and multiple linear regression for prediction in neuroscience. *Journal of Neuroscience Methods* 220(1): 85–91.

Stamps AE (2009) Effects of permeability on perceived enclosure and spaciousness. *Environment and Behavior* 42(6): 864–886.

Sternberg E (2000) An integrative theory of urban design. *Journal of the American Planning Association* 66(3): 265–278.

Sung H, Lee S and Cheon S (2015) Operationalizing Jane Jacobs's urban design theory: Empirical verification from the great city of Seoul, Korea. *Journal of Planning Education and Research* 35(2): 117–130.

Tang J and Long Y (2019) Measuring visual quality of street space and its temporal variation: Methodology and its application in the Hutong area in Beijing. *Landscape and Urban Planning* 191: 103436.

Tucker C, Ostwald M, Chalup S, et al. (2005) A method for the visual analysis of the streetscape. In 5th International Space Syntax Symposium Proceedings, Volume 2, Delft, the Netherlands 13–17 June, 2005. pp. 519–529. Amsterdam: Techne Press.

Xia C, Yeh AG-O and Zhang A (2020) Analyzing spatial relationships between urban land use intensity and urban vitality at street block level: A case study of five Chinese megacities. *Landscape and Urban Planning* 193: 103669.

Xu F, Zhen F, Qin X, et al. (2019) From central place to central flow theory: An exploration of urban catering. *Tourism Geographies* 21(1): 121–142.

Ye Y, Zeng W, Shen Q, et al. (2019) The visual quality of streets: A human-centred continuous measurement based on machine learning algorithms and street view images. *Environment and Planning B-Urban Analytics and City Science* 46(8): 1439–1457.

Yoshimura Y, Kumakoshi Y, Milardo S, et al. (2021) Revisiting Jane Jacobs: Quantifying urban diversity. *Environment and Planning B: Urban Analytics and City Science* 49(4): 1228–1244.

Zhang F, Zhou B, Liu L, et al. (2018) Measuring human perceptions of a large-scale urban region using machine learning. *Landscape and Urban Planning* 180: 148–160.

Enjia Zhang is a Ph.D. candidate in the School of Architecture, Tsinghua University, China. Her research focuses on Data Augmented Design and quantitative urban studies, emphasizing the application of data in urban planning and design.

Hanting Xie was a visiting student of the School of Architecture, Tsinghua University, and now is a Landscape Designer at AECOM San Francisco Office. Before that, she graduated from the University of Pennsylvania with a master's degree in Urban Spatial Analytics and Landscape Architecture. The multi-disciplinary background stimulated her research interests in embedding quantitative analysis into traditional urban studies. She is also interested in the practices of data-driven design in various urban contexts.

Ying Long, Ph.D., is now working in the School of Architecture, Tsinghua University, China as an associate professor. His research focuses on urban science, including applied urban modeling, big data analytics and visualization, quantitative urban studies, planning support systems, Data Augmented Design, and future cities. Dr. Long is also the founder of Beijing City Lab (BCL www.beijingcitylab.com), an open research network for quantitative urban studies. More information is available at http://www.beijingcitylab.com/longy.