

MIT Open Access Articles

Investigating the association between streetscapes and human walking activities using Google Street View and human trajectory data

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Li, Xiaojiang, Santi, Paolo, Courtney, Theodore K., Verma, Santosh K. and Ratti, Carlo. 2018. "Investigating the association between streetscapes and human walking activities using Google Street View and human trajectory data." Transactions in GIS, 22 (4).

As Published: <http://dx.doi.org/10.1111/tgis.12472>

Publisher: Wiley

Persistent URL: <https://hdl.handle.net/1721.1/140888>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



<Running Head> Association between streetscapes and human walking activities

Investigating the association between streetscapes and human walking activities using Google Street View and human trajectory data

Xiaojiang Li¹, Paolo Santi^{1,2}, Theodore K. Courtney^{1,3}, Santosh K. Verma⁴, Carlo Ratti¹

¹Massachusetts Institute of Technology, Cambridge, MA, USA

²Istituto di Informatica e Telematica del Consiglio Nazionale delle Ricerche, Pisa, Italy

³Harvard T. H. Chan School of Public Health, Boston, MA, USA

⁴University of Massachusetts Medical School, Worcester, MA, USA

Correspondence

Xiaojiang Li, Senseable City Laboratory, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

Email: lixiaojiang.gis@gmail.com

Abstract

Having an active lifestyle is recognized to positively contribute to public health. Creating more walkable streets and neighborhoods is an important way to promote an active lifestyle for urban residents. It is therefore important to understand how the urban built environment can influence human walking activities. In this study, we investigated the interaction of human walking activities and physical characteristics of streetscapes in Boston. A large number of anonymous pedestrian trajectories collected from a smartphone application were used to estimate human walking activities. Publicly accessible Google Street View images were used to estimate the amount of street greenery and the enclosure of street canyons, both of which were used to indicate the physical characteristics of streetscapes. The Walk Score and population were also added in the statistical analyses to control the influence of nearby urban facilities and population on human walking activities. Statistical analysis results show that both the street greenery and the enclosure of the street canyons are significantly associated with human walking activities. The associations between the streetscape variables and human walking activities vary in different land use types. The results of this study have implications for designing walkable and healthy

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/tgis.12472](https://doi.org/10.1111/tgis.12472)

This article is protected by copyright. All rights reserved

cities.

1 INTRODUCTION

Physical inactivity increases the risk of cardiovascular disease, certain cancers, hypertension, and obesity (U.S. Department of Health and Human Services, 1996). Lee et al. (2012) estimate that physical inactivity causes 6 to 9% of all deaths from non-communicable diseases worldwide, with the problem of physical inactivity being more prevalent in developed countries (Rundle & Heymsfield, 2016). On the other hand, studies have shown that improving the walkability of streets and neighborhoods helps to promote active lifestyles amongst residents (Duncan, Aldstadt, Whalen, & Melly, 2013; Rundle et al., 2015). Designing more walkable and pedestrian-friendly streets is therefore a promising method to promote human physical activities and prevent chronic health issues (Ewing & Handy, 2009; Yin & Wang, 2016). The relationship between the urban built environment and physical activity is still not exactly quantified and measured, with only some recent efforts in this area (Christian et al., 2011; Zuniga-Teran et al., 2017).

[AQ1: Rundle et al. (2015) cited in the text but not listed in the references. Please give details]

As a basic unit in cities for human activities, streets play an important role in influencing social interactions and affecting people's physical activities and social well-being (Miller & Tolle, 2016; Li, Ratti, & Seiferling, 2017). It is therefore important to understand how the streetscape environment can influence human physical activities. Current literature focuses on smaller neighborhood scale studies (Harvey, Aultman-Hall, Hurley, & Troy, 2015; Yin & Wang, 2016), while considering only reported activities or small-scale samples instead of actual human walking activities (Lee & Li, 2014; Villeneuve et al., 2017). Past studies that rely on small samples or small-scale questionnaires cannot fully represent the entire range of human activities. Collecting street-level built environment data is another challenge to study the connection between the streetscape features and human walking activities (Harvey et al., 2015). Built environment metrics were usually calculated at aggregated areal level to indicate the physical environment of neighborhoods, which cannot fully reflect the built environment at the street level.

[AQ2: Miller and Tolle (2016) cited in the text but not listed in the references. Please give details]

In the mobile and big data era, human trajectory data and street-level images are more

abundant and available, making it possible to study and validate the relationship between actual human activities and streetscape characteristics at a large scale. In this study, we proposed to combine a large number of human trajectory data and Google Street View (GSV) images to investigate the connection between human walking activities and urban built environment at the street level. A large and passively collected human trajectory dataset from a smartphone application was used to estimate the actual human walking activities in Boston. In order to better represent the physical environment at the street level, we used GSV images to measure the streetscape characteristics. Since GSV images were captured along streets with a similar view angle to pedestrians (Li et al., 2015), the built environment metrics derived from GSV images would help to represent the streetscape more objectively.

2 LITERATURE REVIEW

The social-ecological theory of human behavior suggests that environmental factors in cities influence the likelihood of people being physically active (Sallis, Bauman, & Pratt, 1998; Sallis, Floyd, Rodriguez, & Saelens, 2012; Kraus et al., 2015; Zuniga-Teran et al., 2017). Many environmental factors, such as higher housing density, easier access to transit, and greater land use mix, have been found to be influential in determining people's physical activities (Frank, Saelens, Powell, & Chapman, 2007; Leslie et al., 2007; Coogan et al., 2009; Kerr et al., 2014; Zuniga-Teran et al., 2017). The connection between human physical activities and environmental factors gives a motivation to the development of the Walk Score, which incorporates built environment variables together with some other variables of amenity categories such as grocery stores, restaurants, parks, banks, schools, movie theaters, libraries, and other urban facilities to indicate a neighborhood's capacity to support physical activity (Carr, Dunsiger, & Marcus, 2010; Duncan, Aldstadt, Whalen, Melly, & Gortmaker, 2011; Walk Score, 2017). As a publicly available online metric, the Walk Score provides an objective measure of walkability at urban scale (Duncan et al., 2014; Chiu et al., 2015; Gilderbloom, Riggs, & Meares, 2015). However, it is important to observe that the Walk Score measures the number of factors that facilitate walking (e.g. for commuting or going to a store), but not the actual human walking activities. In addition, the Walk Score is only suitable for representing the urban form at the neighborhood or city scale, because streetscape characteristics, which reflect urban features at the street level, are not considered in the algorithm for computing the Walk Score. However, streets carry most

human walking activities in cities (Li et al., 2017), and it is believed that streetscape characteristics would influence human walking preferences directly (Ewing et al., 2005; Harvey et al., 2015). Investigating the interplay between the physical characteristics of streetscapes and human walking activities is therefore needed.

[AQ3: Ewing et al. (2005) cited three times in the text but not listed in the references. Please give details]

Proliferating studies have examined the connection between street-level design qualities and the human perception of environment (Ewing et al., 2005; Asgarzadeh, Lusk, Koga, & Hirate, 2012; Asgarzadeh, Koga, Hirate, Farvid, & Lusk, 2014; Harvey et al., 2015). Ewing et al. (2005) summarized several street design features that are important to pedestrians and the walkability of streets: imageability, enclosure, complexity, and transparency. Harvey et al. (2015) found that streets which are more enclosed by buildings and trees are generally perceived as safer than those streets that are more open and less vegetated. However, the perceived safety is estimated based on the crowdsourcing website, and people's perception of the environment may not be fully represented by those online street-level images. Asgarzadeh et al. (2012, 2014) found that high-rise buildings are more oppressive than low-rise buildings, and street trees would mitigate pedestrians' oppressiveness significantly. The relationship between street greenery and human perceived safety is not consistent in previous studies. Street greenery is believed to have a connection with decreased crime and increased perceived safety (Kuo et al., 2000; Li et al., 2015). However, low street trees that obstruct views are associated with increased occurrence of crime (Donovan & Prestemon, 2012). Tree canopies in streetscapes also contribute to the enclosure and complexity of streetscapes (Arnold, 1993; Jacobs, 1993). On the one hand, the street enclosure contributes to the imageability and local awareness (Lynch, 1960). On the other hand, the enclosed space may cause stress to pedestrians (Asgarzadeh et al., 2014).

[AQ4: Kuo et al. (2000); Lynch (1960) cited in the text but not listed in the references. Please give details]

However, there are still few studies in the current literature investigating the connection between human walking activities and the characteristics of streetscapes at a large scale, while a better understanding of the connection between streetscape and human walking activities would help us design more walkable and healthier cities.

3 METHODOLOGY

3.1 Study area and data collection

As the largest city in Massachusetts, the city of Boston was chosen as the study area. Boston has a land area of 106.7 km² and a total population of 670,000 (as of 2016). Owing to its relatively compact layout, Boston is considered one of the most walkable cities in the United States.

The datasets used in this study include anonymous human trajectory data, GSV data, Walk Score data, land use map, and OpenStreetMap data. The anonymous human trajectory data were collected from an activity-oriented mobile phone application, which was free and downloadable from the App stores. The application was a single proprietary activity-tracking software (health and fitness type), intended for later-model smartphones (Apple iPhone, Android platform devices). Users who downloaded the software presumably had the intention to better track and understand their own activities, as is the case with most available “fitness tracker” or “quantified self”-type software. This application was always on and did not require the user to trigger it. The anonymized data, which includes about 300,000 trips of over 6,000 anonymous users from May 2014–May 2015, records GPS locations and walking behaviors of anonymous individuals in Boston metropolitan area. Although the collected trajectory data may not necessarily be representative of the general population, this anonymized data would give us a new way to understand the connection between human walking activities and the urban built environment at the street level, considering the popularity of the mobile phone application and the massively collected trajectories.

The GSV data was used to measure and estimate the geometries of street canyons and the amount of street greenery. Since GSV panoramas are distributed discretely along streets, we first created samples every 100 m along streets in the study area (Figure 1). Based on those created samples, we further downloaded 10,846 GSV panoramas through the Google Street View API (Google, 2016). The land use map in the study area was taken from MassGIS (2005). We aggregated similar land use types in the original land use map into four major land use types: residential land, commercial land, recreational land, and industrial land. Table 1 shows descriptions of these four aggregated land use types.

3.2 Estimating human activities at street level

The anonymous human trajectory data were used to study human activities in Boston (Vanky, Courtney, Verma, & Ratti, 2016). In the anonymous human trajectory data, a random distance of 0–100 m was removed from the start and end of each trip to further anonymize the users' frequently visited locations (Vanky et al., 2016). However, the original GPS locations in those trajectories are very noisy and have location errors because of the obstruction of the GPS signal by the high-rise building blocks and street trees in cities (Mooney et al., 2016). Figure 2 shows the trajectories of four anonymous individuals in the study area. There are obvious mismatches between the human trajectories and the street maps. In order to correct those trajectories, we used a map-matching algorithm based on the OpenStreetMap. We used the Hidden Markov Map-Matching algorithm (HMM) to match the measured longitudes/latitudes in human trajectory records to roads. The HMM algorithm accounts for the GPS noise and the layout of the road network, and matches the GPS locations to corresponding streets with very good accuracy (Newson & Krumm, 2009). Figure 2 shows the original raw trajectories (purple lines) and the matched trajectories (green lines) of four anonymous individuals. In this study, more than 90% of the raw trajectories were matched successfully to road networks based on the map-matching algorithm. However, some matched trajectories failed to adequately capture the actual paths. To remove the poorest matched trajectories, we calculated the distance offsets between the original trajectories and the matched trajectories, and disregarded those considered outliers using the inter-quartile range.

The matched trajectories were further bound to the study area and aggregated at street level to estimate human walking activities. In order to match with the street greenery variables, we only selected those human trajectories in green seasons (June, July, August, September, and October). In addition, those trajectories on highways, ramps, and motorways were removed from the analysis.

3.3 Streetscape variables

Previous studies have shown that the visibility of greenery and the enclosure of street canyons are associated with human perceptions of the environment (Asgarzadeh et al., 2014; Harvey et al., 2015; Li et al., 2015) and the walkability of the streets (Yin & Wang, 2016). In this study, we used the sky view factor (SVF) and Green View Index (GVI) to represent the enclosure of street

canyons and the amount of street greenery, respectively.

The SVF quantifies the degree of sky visibility or the openness of street canyons. Within street canyons, there are two types of obstructions influencing the enclosure of streetscapes: building blocks and street tree canopies. In this study, we calculated the contribution of these two types of obstructions to the enclosure of street canyons based on GSV panoramas and the building height model. Hemispherical images created from GSV panoramas were used to measure the enclosure of street canyons with consideration of the obstruction of both building blocks and street tree canopies (Li et al., 2017). The ray-tracing algorithm of the building height model, which considers the obstruction effect of buildings only, was used to estimate the street enclosure by buildings. The difference between these two methods defines the contribution of the tree canopies to the enclosure of street canyons.

In this study, we used the method proposed by Li et al. (2017) to classify the hemispherical images into three major types: buildings, tree canopies, and sky pixels. Figure 3 shows hemispherical images generated based on GSV (Figure 3b) and the ray-tracing algorithm of the building height model (Figure 3c) at one site of the study area. Figure 3d shows the classification result of a GSV-based hemispherical image.

The SVF can then be calculated based on the sky classification in the generated hemispherical images using the photographic method. The photographic method (Steyn, 1980) first divides the fisheye image into n concentric annular rings of equal width, and then sums up all annular sections representing the visible sky. The SVF is then calculated as

$$SVF = \frac{1}{2\pi} \sin\left(\frac{\pi}{2n}\right) \sum_{i=1}^n \sin\left(\frac{\pi(2i-1)}{2n}\right) \alpha_i \quad (1)$$

where n is the number of rings, i is the ring index, and α_i is the angular width in the i th ring. The SVF indicates the openness of the street canyon, and its value ranges from 0 to 1. The SVF value is 1 when there is no obstruction, and 0 when the sky is totally obstructed. Therefore, the enclosure can be calculated by

$$\text{Enclosure} = 1 - SVF \quad (2)$$

The enclosure of street canyons by buildings can be estimated using the same method based on the simulated hemispherical images (Figure 3c) from the building height model. Since the ray-tracing algorithm considers the obstruction of buildings only, enclosure of the street canyons caused by tree canopies can be estimated as the enclosure difference between the GSV method and the ray-tracing method in the building height model.

The GVI, which measures the visibility of street greenery based on GSV images (Li et al., 2015), was used to measure the amount of street greenery within street canyons. In this study, we calculated the GVI using six horizontal static GSV images only, since the vertical structure of the street greenery has already been considered in the street enclosure metrics. The GVI in this study was calculated as:

$$GVI = \frac{\sum_{i=1}^6 Area_{g_i}}{\sum_{i=1}^6 Area_{t_i}} \quad (3)$$

where $Area_{g_i}$ is the number of green pixels in a static GSV image, $Area_{t_i}$ is the number of total pixels in one GSV image. Only those GSV images taken in green seasons will be used in the computation.

Other than the geometry of street canyons and the amount of street greenery, the Walk Score and population were also considered in the analysis. Considering the fact that the Walk Score measures the proximity to nearby urban facilities, we also added the Walk Score in our analysis as a confounding variable. We collected the Walk Score for all created sample sites through the Walk Score API by using the coordinates of those sample sites as input. The population information was derived from the 2009–2014 five-year American Census Survey data at census tract level.

Different representations of land use diversity may impact the association between neighborhood design and specific walking behaviors (Christian et al., 2011). In addition, land use types were also used to represent the built environment, considering that the human activities and the interaction of people with the physical environment would be different in different land use types. Therefore, different land use types were also considered in the analysis.

3.4 Statistical analysis

Four streetscape variables were selected (street enclosure by building, street enclosure by tree, GVI, and enclosure of the streetscapes) in the statistical analysis. The Walk Score and population were also added in order to control the influence of urban amenities and population on the pedestrian trip number. Pearson correlation analysis was first conducted between the pedestrian trip number and the chosen streetscape variables.

In order to further investigate the impact of urban design features on human walking activities, we applied regression models to study the association between the independent variables and the pedestrian trip number. The variable of actual enclosure of street canyons was

not included in the regression analysis, because it has strong correlation with street enclosure by buildings and street enclosure by trees. Different regression models were applied for each type of land use.

The existence of spatial autocorrelation would violate the assumption of the linear regression model (Talen & Anselin, 1998); to account for this, we also checked the spatial autocorrelation of the regression residuals by calculating Moran's I statistics. If there was a significant spatial autocorrelation, we then used the spatial regression models to study the associations between the dependent variable and the independent variables in different land use types (Anselin & Bera, 1998; Anselin, 2005). There have been two common approaches to include the spatial dependence in spatial regression models—spatial error regression model (SAR_{err}) and spatial lagged regression model (SAR_{lag})—which incorporate the spatial autocorrelation effects in the residual error term and the dependent variable, respectively. The Lagrange multiplier and robust Lagrange multiplier tests were used to help choose the right type of spatial regression model.

4 RESULTS

Figure 4a shows the spatial distribution of the human walking activities in terms of trip number at street level in the study area. An expected activity pattern can be observed, with the downtown and Back Bay areas having more intensive human activities than the peripheral part of the study area. This is consistent with the fact that the downtown area has the most commercial shops, workplaces, and public transportation stops. In order to make the street-level trip number directly comparable with other independent variables, we further overlaid the sample sites on the street-level trip number map to get the trip number map at site level (Figure 4b).

Figure 5 shows the spatial distributions of the four independent variables. Generally, the Walk Score map (Figure 5a) has a similar spatial distribution to the trip number map (Figure 4). The Walk Score in the downtown area has larger values than the southern and southwestern areas. This is because the downtown area has more urban facilities, which would further have larger Walk Score values, considering the fact that the Walk Score is calculated based on the proximity to different urban facilities and the density of the urban facilities. In the map of street enclosure by buildings (Figure 5c), the downtown area has higher enclosure level compared with the southern part of the study area. This is explained by the fact that buildings in the southern part of

the study area are much lower compared with the downtown areas. The GVI map (Figure 5b) and the street enclosure by street trees map (Figure 5d) have a similar spatial distribution. In both of these maps, the peripheral parts have higher values than the central part of the study area. This is because both variables reflect the amount of street greenery, and the peripheral parts have more street greenery than the central part of the study area. The difference between these two variables is that the GVI represents the horizontal visibility of street greenery, while the street enclosure by trees indicates the amount of street greenery overhead.

Table 2 shows the Pearson correlation coefficients between the trip number and independent variables. The trip number has a significant and positive correlation with the Walk Score. The population is significantly and negatively correlated with the trip number. However, the correlation coefficient is very low. The street enclosure by buildings has a very significant and positive correlation with the trip number. Both the GVI and the street enclosure by trees have a significant and negative correlation with the trip number. There is a weakly significant and positive correlation between the trip number and the enclosure of the streetscape, which is enclosed by joint building blocks and street greenery.

Considering the fact that the interaction of pedestrians and streetscape variables is different in different land use types, we further investigated the relationship between the trip number and the independent variables in different land use types. We used ordinary least squares (OLS) regression models to investigate the associations between the trip number and different independent variables. Since the trip number variable is very skewed, we used a log transform to make it satisfy the assumption of normality. The enclosure of the streetscape has a very significant and strong correlation with the variable of street enclosure by buildings, therefore we only consider the GVI, street enclosure by buildings, street enclosure by trees, population, and Walk Score in the regression models as independent variables.

Table 3 shows the OLS regression analysis results between the trip number and the streetscape variables in different land use types. Generally, the Walk Score has a very positive and significant association with the pedestrian trip number in all four land use types. The population is significantly and negatively associated with the pedestrian trip number. The street enclosure by buildings has a very significant and positive association with the trip number. The GVI and street enclosure by trees have different associations with the trip number in different land use types. In the residential land, commercial land, and recreational land, there is no

significant association between the GVI and the trip number. However, for the industrial land, there is a significant and negative association between the trip number and the GVI. Different from the street enclosure by buildings, the street enclosure by trees has no significant association with the trip number in all land use types, except recreational land. In the recreational land use, there is a significant and negative association between the street enclosure by trees and trip number.

The residuals in the four OLS regression models have significant spatial autocorrelation, therefore we used the spatial regression models to investigate the association between the trip number and independent variables in different land use types. For comparison, spatial regression models with and without confounding variable were applied. Table 4 shows the spatial regression results between the dependent variable and independent variables in four different land use types. Generally, adding the confounding variable into the regression models does not have much influence on the significance of independent variables. For the residential land, the Walk Score remains a significant contributor to the human trip number in the spatial error regression model (SAR_{err}). The population is significantly and negatively associated with the trip number. Different from the OLS model, after controlling the spatial autocorrelation and the confounding variable Walk Score, both the GVI and the street enclosure by buildings have significantly negative associations with the trip number. Similar to the OLS model, the street enclosure by trees has non-significant association with the trip number. For commercial land, the Walk Score has a significant and positive association with the trip number. There is no significant association between the population and the trip number. The GVI is significantly and negatively associated with the trip number. The street enclosure by buildings and the street enclosure by trees both have no significant association with the trip number. For recreational land, the Walk Score and population have significantly positive and negative associations with the trip number, respectively. There is no significant association between the trip number and the other three independent variables after controlling the spatial autocorrelation. For industrial land, the Walk Score and street enclosure by buildings have significant and positive associations with the trip number in the spatial regression model. The GVI and street enclosure by trees both have no significant association with the trip number. The population has a significantly negative association with the trip number.

5 DISCUSSION

This study investigates the relationship between human walking activities and the physical characteristics of streetscapes. Pervasively collected human trajectory data was used to estimate human walking activities in the study area. The large number of human trajectories permit a more unbiased estimation of the actual human walking activities compared with previous studies based on small-scale samples. In order to represent the characteristics of streetscapes in the study area, several streetscape variables were calculated from GSV data and the building height model. Tens of thousands of street-level images and panoramas were used to calculate the streetscape variables for sample sites along streets at a fine level. Considering the fact that the Walk Score is a composite measure of the potential for walking with consideration of the density of, and proximity to, urban facilities, the Walk Score was also selected in the analysis as the confounding variable to investigate the connection between the characteristics of streetscapes and human walking activities.

Statistical analysis results show that the associations between human walking activities and the streetscape variables vary among different land use types after controlling the confounding variable of the Walk Score and population. The visibility of the street greenery has different associations with human walking activities among different land use types. In residential and commercial land use areas, the visibility of the street greenery is negatively associated with human walking activities. For recreational land and industrial land, there is no significant association between the visibility of the street greenery and human walking activities. Different from previous studies, which showed that the enclosure of the streetscape contributes to increased walking activities, this study finds that the street enclosure has different associations with human walking activities among different land use types. The street enclosure by buildings and the street enclosure by trees have different associations with human walking activities in the study. The street enclosure by trees has a significant and negative correlation with human walking activities. However, regression models show that the street enclosure by trees has no significant association with human walking activities in all land use types after controlling other independent variables and the spatial autocorrelation. The street enclosure by buildings would give more power to the variance of the dependent variable (the trip number), especially in for residential land and industrial land. The findings provide a reference for physical activity promotion intervention programs in cities.

This study contributes a new methodology to investigating the associations between the urban built environment and human walking activities at the street level by combining street-level images and human trajectory data. The human trajectory data and street-level images help to objectively represent human walking activities and streetscapes, respectively. Therefore, the combination of these two types of dataset would help us to better understand the connection between human activities and the built environment. With the public and global availability of street-level images and the increasingly abundant human GPS trajectory data, it is possible to deploy the proposed workflow to other cities to understand the connections between streetscapes and human activities, which would further benefit urban design and planning in those cities. In addition, the proposed methodology using publicly accessible and globally available street-level images provides us with a new tool to test the social–ecological theory which could promote public health research.

Additionally, the connections between the streetscape metrics and human walking activities could support the development of better metrics to measure the walkability in future. Statistical analyses results show that the current state-of-the-art walkability metric, Walk Score, shows some power at indicating the walkability at city scale, but still does not fully represent actual human walking activities. More factors, especially the streetscape variables, need to be considered to better represent the real walkability at street level.

Although this study used a large number of pervasively collected anonymous trajectories to study human walking activities and investigated the association between the physical characteristics of streetscapes and human walking activities, the current study still has some limitations. First, the large number of human trajectories still cannot fully represent all human walking behaviors in the study area. The observed population may not necessarily be representative of the general population. Demographic details that could be desirable were redacted to protect user anonymity. Generally, the population of smartphone owners and mobile application users tends to skew toward younger, more affluent individuals, though without demographic information this could not be confirmed. If we presume that these users were more affluent and younger, then the results may underestimate the behavior of older, less affluent individuals, who could have different mobility patterns. Big or “found” data such as those utilized here are a new territory for research. While offering information at large scale, they may lack detailed information typically collected in traditional studies. Future studies should try to

combine different sources, including passively collected trajectory data to get more objective estimations of human walking activities.

Second, although the map-matching algorithm can match most trajectories successfully, a small number of paths failed to accurately capture the path represented by the GPS coordinates. These mismatched trajectories would misrepresent the actual human walking activities at street level, which could further bring noise into the statistical analyses. In addition, walking preferences between local residents and tourists could be different. Future studies should also consider the difference between tourists and local residents.

There are many other streetscape features that could influence human walking activities. In this study, we only considered the visibility of the street greenery and the enclosure of street canyons using the Walk Score as the confounding variable. Future studies should consider more variables in the analysis of the connection between the streetscape characteristics and human walking activities.

6 CONCLUSION

This study investigated the actual human activities using large-scale passively collected human trajectory data and studied the connection between streetscape characteristics and human walking activities. Streetscape characteristics have different associations with human walking activities in different land use types. The enclosure of street canyons is an important factor associated with human walking activities at street level. The amount of vegetation has different associations with walking activities among different land use types. This study provides a meaningful reference for urban planners and designers seeking to create more walkable streets and healthier cities. This study also demonstrates the usefulness of passive, pervasive mobile devices and publicly accessible street-level images in evaluating urban space and investigating the interplay between the urban built environment and human activities.

Future studies should also investigate the connection between streetscape characteristics and human walking activities in different cities of different climate zones. The different roles of enclosure by buildings and trees should also be investigated in the future, considering the fact that enclosure by trees could play an important role in influencing human thermal comfort by providing shade during hot summers.

ACKNOWLEDGMENTS

X. Li, P. Santi, and C. Ratti would like to gratefully thank: UBER, Allianz, AIT, the Amsterdam Institute for Advanced Metropolitan Solutions, Ericsson, the Fraunhofer Institute, Liberty Mutual Institute, SNCF, the Kuwait-MIT Center for Natural Resources and the Environment, Singapore-MIT Alliance for Research and Technology (SMART), and all the members of the MIT Senseable City Lab Consortium for supporting this research.

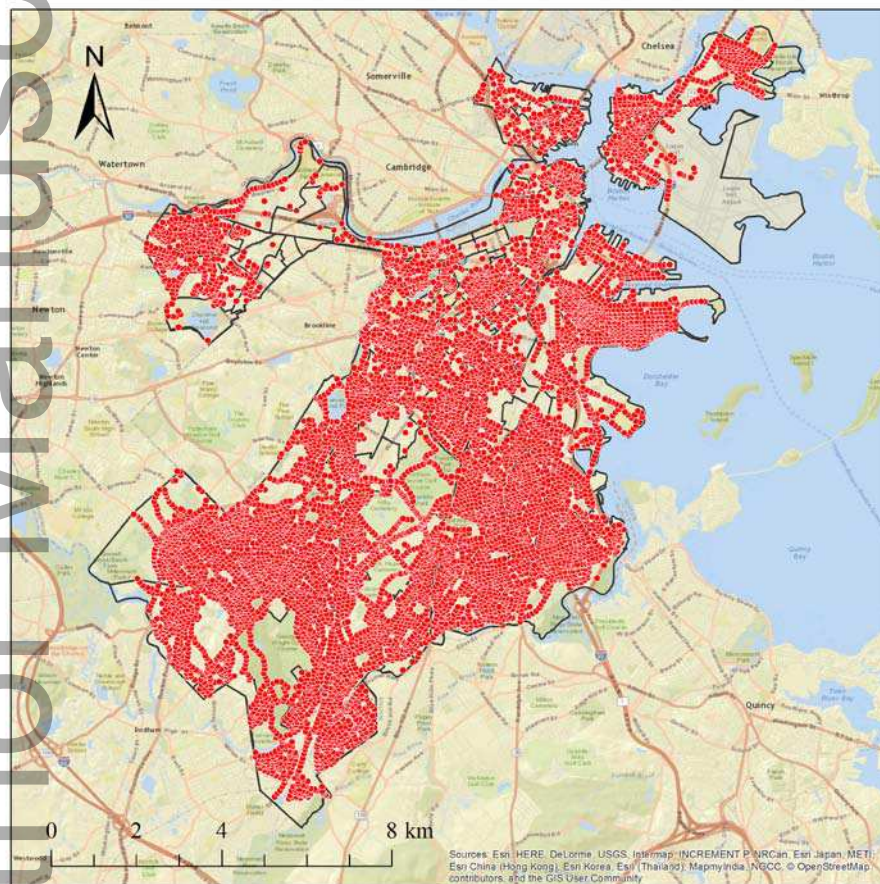


FIGURE 1 The location of the study area and the created sample sites



FIGURE 2 Map matching of human trajectories to streets: the purple lines represent the raw trajectories of anonymous individuals, and the green lines represent the matched trajectories based on OpenStreetMap

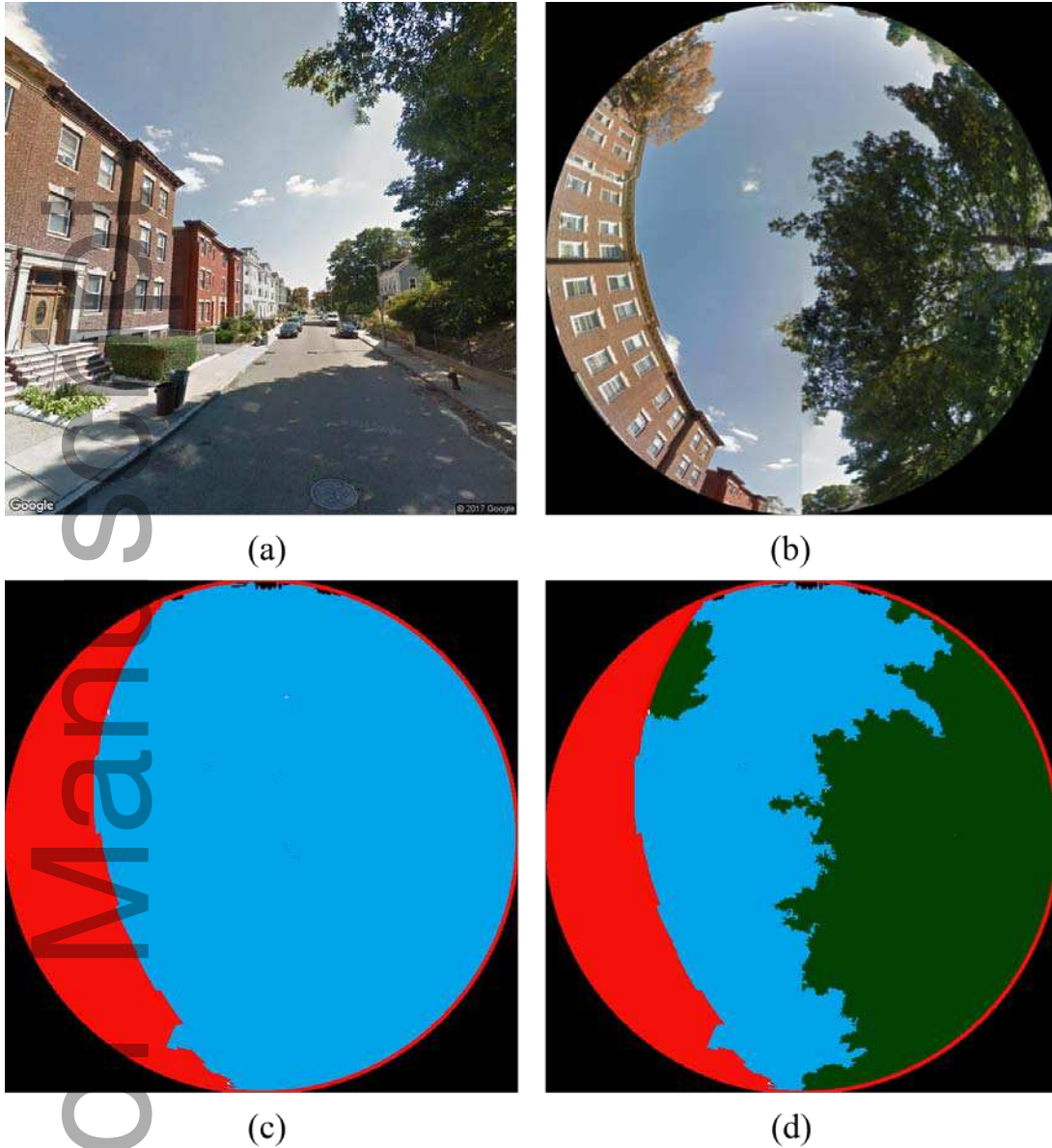


FIGURE 3 Estimating the enclosure of street canyons: (a) profile view of streetscape at one site; (b) hemispherical image generated from GSV panorama; (c) synthetic open sky image from the building height model; (d) classification result of the generated hemispherical image. (Red represents the building, cyan represents the sky, and green represents the vegetation)

FIGURE 4 The spatial distributions of the trip number at street level and site level in Boston

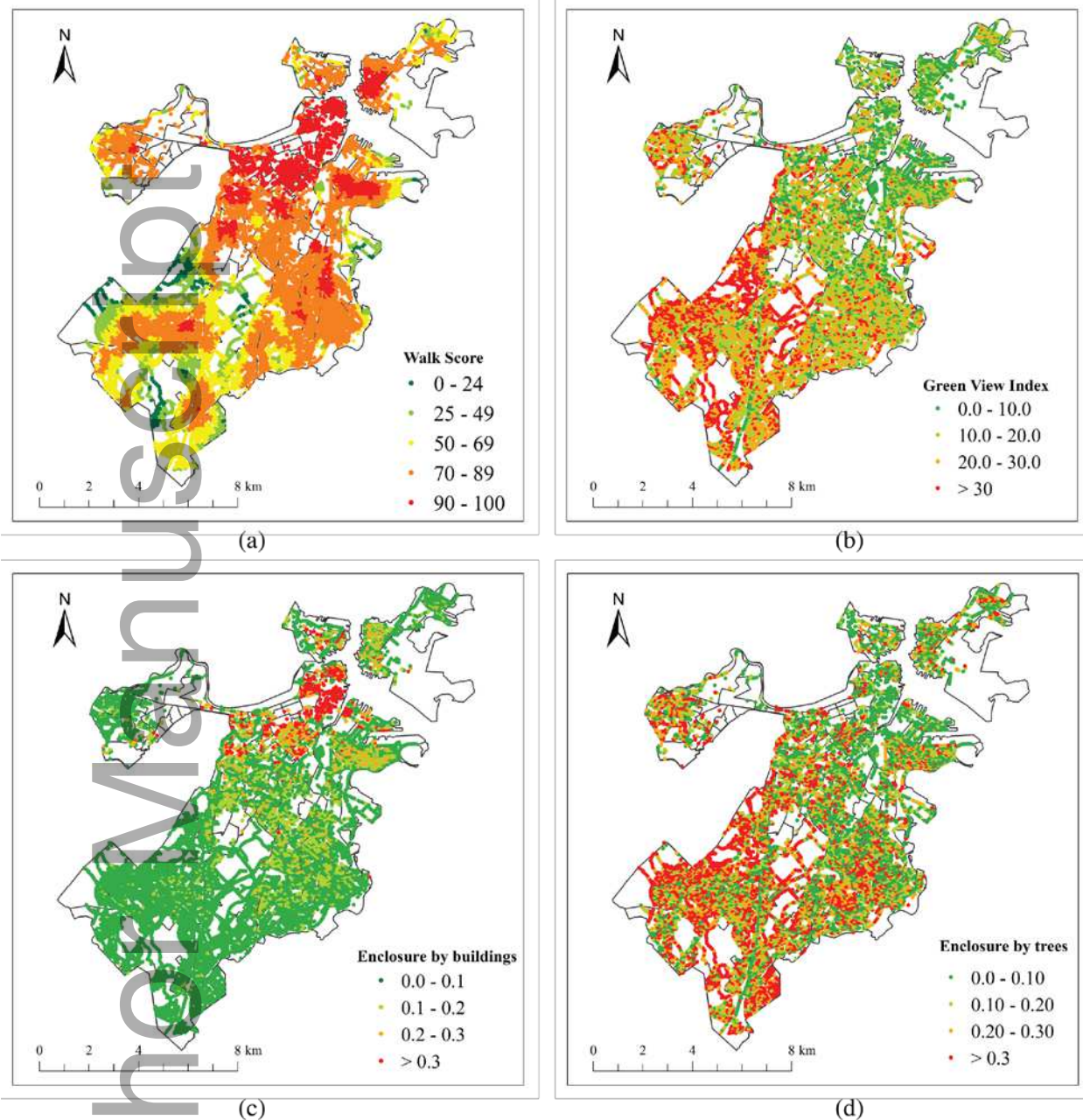


FIGURE 5 The spatial distributions of independent variables: (a) the Walk Score; (b) the GVI; (c) the enclosure of streetscapes enclosed by buildings; (d) the enclosure of streetscapes enclosed by street tree canopies

TABLE 1 Aggregated land use types in Boston and the description of each land use type

Land use types	Descriptions
Residential land	High-density residential land, medium-density residential land, multifamily residential land.
Commercial land	Malls, shopping centers and larger strip commercial areas, plus neighborhood stores and medical offices (not hospitals).
Recreational land	Lands comprising open land, institutional facilities, wetlands, marina, pasture, public open green spaces, and cropland.
Industrial land	Light and heavy industry, including buildings, equipment, and parking areas; transportation land; mining.

TABLE 2 Correlations between trip number and independent variables

Correlation analysis	Pearson's correlation	Sig. (2-tailed)	N
Walk Score	0.27**	0.000	10,846
Population	-0.04**	0.000	
GVI	-0.19**	0.000	
Enclosure of streetscape	0.09**	0.000	
Street enclosure by buildings	0.49**	0.000	
Street enclosure by trees	-0.13**	0.000	

**Correlation significant at the 0.01 level (2-tailed).

TABLE 3 OLS regression models in different land use types

Land use	Variables	Coefficients	z Values	
Residential land (N=6,813)	Walk Score	0.04**	23.95	
	Population	-0.17×10^{-3} **	-12.57	
	GVI	-0.6×10^{-2}	-2.22	Adj. R square: 0.30
	Enclosure by buildings	7.26**	22.93	F statistic: 596.36
	Enclosure by trees	0.22	1.91	Moran's I of residuals: 0.30**
Commercial land (N=1,251)	Walk Score	0.05**	11.58	
	Population	-0.16×10^{-3} **	-5.17	
	GVI	-1.20×10^{-2}	-1.81	Adj. R square: 0.50
	Enclosure by buildings	6.82**	21.98	F statistic: 250.0
	Enclosure by trees	0.27	0.77	Moran's I of residuals: 0.32**
Recreational land (N=1,860)	Walk Score	0.03**	11.28	
	Population	-0.22×10^{-3} **	-10.59	
	GVI	0.78×10^{-2}	1.64	Adj. R square: 0.23
	Enclosure by buildings	5.72**	11.98	F statistic: 114.35
	Enclosure by trees	-0.67 **	-3.07	Moran's I of residuals: 0.41**
Industrial land (N=922)	Walk Score	0.04**	11.27	
	Population	-0.3×10^{-3} **	-10.22	
	GVI	-2.67×10^{-2} **	-3.84	Adj. R square: 0.29
	Enclosure by buildings	4.73**	8.73	F statistic: 74.85
	Enclosure by trees	0.30	0.80	Moran's I of residuals: 0.32**

**Significant at the 0.01 level (2-tailed).

TABLE 4 Spatial error regression (SAR_{err}) models in different land use types

Land use types	Variables	Models with and without confounding variable	
		Coefficients (z values)	Coefficients (z values)
Residential land (N=6,813)	Walk Score	0.04** (20.01)	—
	Population	-0.08×10^{-3} ** (-5.65)	-0.08×10^{-3} ** (-5.05)
	GVI	-1.50×10^{-2} ** (-6.44)	-1.94×10^{-2} ** (-8.17)
	Enclosure by buildings	-1.17** (-3.77)	-0.33 (-1.03)
	Enclosure by trees	0.02 (0.26)	-0.07 (-0.79)
	Spatial error term	0.97** (154.8)	0.98** (171.0)
	Adj. R square	0.60	0.58
Akaike information criterion	23,182.5	23,569.2	
Commercial land (N=1,251)	Walk Score	0.03** (4.47)	—
	Population	0.02×10^{-3} (0.60)	0.02×10^{-3} (0.63)
	GVI	-1.35×10^{-2} ** (-2.80)	-1.63×10^{-2} ** (-3.42)
	Enclosure by buildings	0.41 (1.16)	0.49 (1.38)
	Enclosure by trees	0.14 (0.55)	0.13 (0.51)
	Spatial error term	0.90** (56.29)	0.91** (62.33)
	Adj. R square	0.75	0.75
Akaike information criterion	4,070.2	4,087.4	
Recreational land (N=1,860)	Walk Score	0.02** (5.01)	—
	Population	-0.11×10^{-3} ** (-5.69)	-0.10×10^{-3} ** (-4.89)

	GVI	0.06×10^{-2} (0.16)	0.18×10^{-2} (0.50)
	Enclosure by buildings	0.23 (0.55)	0.31 (0.75)
	Enclosure by trees	-0.24 (-1.46)	-0.25 (-1.50)
	Spatial error term	0.89** (58.10)	0.90** (63.22)
Adj. R square		0.61	0.61
Akaike information criterion		6,292.5	6,313.0
Industrial land (N=922)	Walk Score	0.02** (7.56)	—
	Population	-0.08×10^{-3} ** (-3.42)	-0.05×10^{-3} (-1.56)
	GVI	-0.38×10^{-2} (-0.70)	-0.03×10^{-2} (-0.06)
	Enclosure by buildings	1.91** (4.33)	2.45** (5.34)
	Enclosure by trees	0.27 (1.00)	0.40 (1.43)
	Spatial error term	0.90** (45.89)	0.90** (45.89)
Adj. R square		0.61	0.56
Akaike information criterion		3,134.3	3,222.6

**Significant at the 0.01 level (2-tailed).

References

- Anselin, L. (2005). *Exploring spatial data with GeoDaTM: A workbook*. Champaign-Urbana, IL: Spatial Analysis Laboratory, Department of Agricultural and Consumer Economics, University of Illinois.
- Anselin, L., & Bera, A. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In A. Ullah & D. E. A. Giles (Eds.), *Handbook of applied economic statistics* (pp. 237–289). New York, NY: Marcel Dekker.
- Arnold, H. F. (1993). *Trees in urban design*. New York, NY: Van Nostrand Reinhold.
- Asgarzadeh, M., Koga, T., Hirate, K., Farvid, M., & Lusk, A. (2014). Investigating oppressiveness and spaciousness in relation to building, trees, sky and ground surface: A study in Tokyo. *Landscape & Urban Planning, 131*, 36–41.
- Asgarzadeh, M., Lusk, A., Koga, T., & Hirate, K. (2012). Measuring oppressiveness of streetscapes. *Landscape & Urban Planning, 107*(1), 1–11.
- Carr, L. J., Dunsiger, S. I., & Marcus, B. H. (2010). Validation of Walk Score for estimating access to walkable amenities. *British Journal of Sports Medicine, 45*(14), 1144–1148.
- Chiu, M., Shah, B. R., Maclagan, L. C., Rezai, M. R., Austin, P. C., & Tu, J. V. (2015). Walk Score[®] and the prevalence of utilitarian walking and obesity among Ontario adults: A cross-sectional study. *Health Reports, 26*(7), 3.
- Christian, H. E., Bull, F. C., Middleton, N. J., Knuiaman, M. W., Divitini, M. L., Hooper, P., ..., & Giles-Corti, B. (2011). How important is the land use mix measure in understanding walking behaviour? Results from the RESIDE study. *International Journal of Behavioral Nutrition & Physical Activity, 8*(1), 55.
- Coogan, P. F., White, L. F., Adler, T. J., Hathaway, K. M., Palmer, J. R., & Rosenberg, L. (2009). Prospective study of urban form and physical activity in the Black Women's Health Study. *American Journal of Epidemiology, 170*(9), 1105–1117.
- Dodge, S., Weibel, R., & Forootan, E. (2009). Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects. *Computers, Environment & Urban Systems, 33*(6), 419–434.

[AQ5: Dodge, Weibel and Forootan (2009) not cited in the text. Please add a text citation or delete the reference]

- Donovan, G. H., & Prestemon, J. P. (2012). The effect of trees on crime in Portland, Oregon. *Environment & Behavior, 44*(1), 3–30.
- Duncan, D. T., Aldstadt, J., Whalen, J., & Melly, S. J. (2013). Validation of walk scores and transit scores for estimating neighborhood walkability and transit availability: A small-area analysis. *GeoJournal, 78*(2), 407–416.
- Duncan, D. T., Aldstadt, J., Whalen, J., Melly, S. J., & Gortmaker, S. L. (2011). Validation of Walk Score® for estimating neighborhood walkability: An analysis of four US metropolitan areas. *International Journal of Environmental Research & Public Health, 8*(11), 4160–4179.
- Duncan, D. T., Sharifi, M., Melly, S. J., Marshall, R., Sequist, T. D., Rifas-Shiman, S. L., & Taveras, E. M. (2014). Characteristics of walkable built environments and BMI z-scores in children: Evidence from a large electronic health record database. *Environmental Health Perspectives, 122*(12), 1359.
- Ewing, R., & Handy, S. L. (2009). Measuring the unmeasurable: Urban design qualities related to walkability. *Journal of Urban Design, 14*, 65–84.
- Frank, L. D., Saelens, B. E., Powell, K. E., & Chapman, J. E. (2007). Stepping towards causation: Do built environments or neighborhood and travel preferences explain physical activity, driving, and obesity? *Social Science & Medicine, 65*(9), 1898–1914.
- Gilderbloom, J. I., Riggs, W. W., & Meares, W. L. (2015). Does walkability matter? An examination of walkability's impact on housing values, foreclosures and crime. *Cities, 42*, 13–24.
- Google. (2016). *Google Street View Image API*. Retrieved from <https://developers.google.com/maps/documentation/streetview/>
- Harvey, C., Aultman-Hall, L., Hurley, S. E., & Troy, A. (2015). Effects of skeletal streetscape design on perceived safety. *Landscape & Urban Planning, 142*, 18–28.
- Jacobs, A. B. (1993). *Great streets*. Cambridge, MA: MIT Press.
- Kerr, J., Norman, G., Millstein, R., Adams, M. A., Morgan, C., Langer, R. D., & Allison, M. (2014). Neighborhood environment and physical activity among older women: Findings from the

San Diego Cohort of the Women's Health Initiative. *Journal of Physical Activity and Health*, 11(6), 1070–1077.

Kraus, W. E., Bittner, V., Appel, L., Blair, S. N., Church, T., Després, J. P., ..., & Whitsel, L. (2015).

American Heart Association Physical Activity Committee of the Council on Lifestyle and Metabolic Health, Council on Clinical Cardiology, Council on Hypertension, and Council on Cardiovascular and Stroke Nursing: The National Physical Activity Plan, a call to action from the American Heart Association, and a science advisory from the American Heart Association. *Circulation*, 131(21), 1932–1940.

Lee, C., & Li, L. (2014). Demographic, physical activity, and route characteristics related to school transportation: An exploratory study. *American Journal of Health Promotion*, 28(3-Suppl.), S77–S88.

Lee, I. M., Shiroma, E. J., Lobelo, F., Puska, P., Blair, S. N., Katzmarzyk, P. T., & Lancet Physical Activity Series Working Group. (2012). Effect of physical inactivity on major non-communicable diseases worldwide: An analysis of burden of disease and life expectancy. *The Lancet*, 380(9838), 219–229.

Leslie, E., Coffee, N., Frank, L., Owen, N., Bauman, A., & Hugo, G. (2007). Walkability of local communities: Using geographic information systems to objectively assess relevant environmental attributes. *Health & Place*, 13(1), 111–122.

Li, X., Ratti, C., & Seiferling, I. (2017). Mapping urban landscapes along streets using Google Street View. In M. P. Peterson (Ed.), *Advances in cartography and GIScience: Selections from the International Cartographic Conference 2017* (pp. 341–356). Berlin, Germany: Springer.

Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., & Zhang, W. (2015). Assessing street-level urban greenery using Google Street View and a modified green view index. *Urban Forestry & Urban Greening*, 14(3), 675–685.

Lo, R. H. (2009). Walkability: What is it? *Journal of Urbanism*, 2, 145–166.

[AQ6: Lo (2009) not cited in the text. Please add a text citation or delete the reference]

MassGIS. (2005). MassGIS – Land use. Retrieved from <http://www.mass.gov/anf/research-and-tech/it-serv519and-support/application-serv/office-of-geographic-information-massgis/datalayers/lus2005.html>

- Mooney, S. J., Sheehan, D. M., Zulaika, G., Rundle, A. G., McGill, K., Behrooz, M. R., & Lovasi, G. S. (2016). Quantifying distance overestimation from global positioning systems in urban spaces. *American Journal of Public Health, 106*(4), 651–653.
- Newson, P., & Krumm, J. (2009). Hidden Markov map matching through noise and sparseness. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (pp. 336–343). Seattle, WA: ACM.
- Rundle, A. G., & Heymsfield, S. B. (2016). Can walkable urban design play a role in reducing the incidence of obesity-related conditions? *Journal of the American Medical Association, 315*(20), 2175–2177.
- Sallis, J., Bauman, A., & Pratt, M. (1998). Environmental and policy interventions to promote physical activity. *American Journal of Preventive Medicine, 15*(4), 379–397.
- Sallis, J. F., Floyd, M. F., Rodriguez, D. A., & Saelens, B. E. (2012). Role of built environments in physical activity, obesity, and cardiovascular disease. *Circulation, 125*(5), 729–737.
- Steyn, D. G. (1980). The calculation of view factors from fisheye lens photographs. *Atmosphere-Ocean, 18*, 254–258.
- Talen, E., & Anselin, L. (1998). Assessing spatial equity: An evaluation of measures of accessibility to public playgrounds. *Environment and Planning A, 30*, 595–613.
- U.S. Department of Health and Human Services. (1996). *Physical activity and health: A report of the Surgeon General*. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion.
- Vanky, A., Courtney, T., Verma, S., & Ratti, C. (2016, July). One to many: Opportunities to understanding collective behaviors in urban environments through individual's passively collected locative data. In N. Streitz & P. Markopoulos (Eds.), *DAPI 2016* (Lecture Notes in Computer Science, Vol. 9749, pp. 482–493). Cham, Switzerland: Springer.
- Villeneuve, P., Root, A., Kumar, N., Ambrose, S., DiMuzio, J., Rainham, D., & Ysseldyk, R. (2017). Are neighbourhood walkability and greenness associated with increased physical activity and better self-rated health? Findings from the Beyond Health Survey of Adults in Ottawa, Canada. *Journal of Transport & Health, 5*, S88.

Walk Score. (2017). *Walk score methodology*. Retrieved from <http://www.walkscore.com/live-more/>

Yin, L., & Wang, Z. (2016). Measuring visual enclosure for street walkability: Using machine learning algorithms and Google Street View imagery. *Applied Geography, 76*, 147–153.

Zuniga-Teran, A. A., Orr, B. J., Gimblett, R. H., Chalfoun, N. V., Marsh, S. E., Guertin, D. P., & Going, S. B. (2017). Designing healthy communities: Testing the walkability model. *Frontiers of Architectural Research, 6*(1), 63–73.

Author Manuscript