


Article

# A Data-Driven Framework for Walkability Measurement with Open Data: A Case Study of Triple Cities, New York

Chengbin Deng <sup>1,\*</sup> , Xiaoyu Dong <sup>1</sup>, Huihai Wang <sup>1</sup>, Weiyong Lin <sup>1,2</sup>, Hao Wen <sup>1</sup>, John Frazier <sup>1</sup>, Hung Chak Ho <sup>3</sup>  and Louisa Holmes <sup>1</sup> 

<sup>1</sup> Department of Geography, State University of New York at Binghamton, Binghamton, NY 13902, USA; xdong6@binghamton.edu (X.D.); hwang164@binghamton.edu (H.W.); wlin56@tamu.edu (W.L.); hwen8@binghamton.edu (H.W.); frazier@binghamton.edu (J.F.); lmholmes@binghamton.edu (L.H.)

<sup>2</sup> Department of Geography, Texas A&M University, College State, TX 77843, USA

<sup>3</sup> Department of Urban Planning and Design, The University of Hong Kong, Hong Kong, China; hcho21@hku.hk

\* Correspondence: cdeng@binghamton.edu

Received: 10 November 2019; Accepted: 30 December 2019; Published: 9 January 2020



**Abstract:** Walking is the most common, environment-friendly, and inexpensive type of physical activity. To perform in-depth walkability analysis, one option is to objectively evaluate different aspects of built environment related to walkability. In this study, we proposed a computational framework for walkability measurement using open data. Three major steps of this framework include the web scrapping of publicly available online data, determining varying weights of variables, and generating a synthetic walkability index. The results suggest three major conclusions. First, the proposed framework provides an explicit mechanism for walkability measurement. Second, the synthetic walkability index from this framework is comparable to Walk Score, and it tends to have a slightly higher sensitivity, especially in highly walkable areas in urban core. Third, this framework was effectively applied in a metropolitan area that contains three small cities that together represent a small, old shrinking region, which extends the topical area in the literature. This framework has the potential to quantify walkability in any city, especially cities with a small population where walkability has rarely been studied, or those having no quantification indicator. For such areas, researchers can calculate the synthetic walkability index based on this framework, to assist urban planners, community leaders, health officials, and policymakers in their practices to improve the walking environment of their communities.

**Keywords:** walkability; GIS; data-driven method; open data; shrinking small cities; geospatial big data; smart cities; urban mapping; urban built environment; urban planning

## 1. Introduction

Walking plays a key role in promoting healthy communities, increasing economic opportunities, and strengthening social connections [1]. It attracts attentions of urban planners, health officials, geographers, social scientists, and policy makers. Besides, walking is the most common form of physical activity among adults [2]. The extent to which adults will walk in their neighborhoods relies on a variety of factors in urban environments [3–5]. As a popular term in the urban studies and public health literature, walkability is increasingly employed to describe the capacity of a community to support its residents' walking activity [6,7]. The characteristics associated with neighborhood walkability have positive impacts on three aspects.

The first is about health benefits. Walking is an accessible form of physical activity that may be readily adopted by people of all ages, requiring little skill, and presenting a limited opportunity for injury [8–11]. This is particularly important in the context of the increasing obesity problem worldwide and in the U.S. [12]. In 2014, the obesity rate of American adults was over 37% compared to a rate of less than 15 percent in 1960 [13]. Obesity carries individual as well as collective costs, and it is a major risk factor for cardiovascular disease, diabetes, depression, sleep apnea, hypertension, bone and joint diseases, and various cancers [10,14]. Its associated annual medical costs reached \$147 billion in 2008 [15]. A lack of physical activity is a primary driver of the growing obesity rate [12]. The World Health Organization reports that 25% of adults have physical activities less than 150 min per week [16]. This deficit led the U.S. Federal Government published its physical activity guidelines to meet Healthy People 2020 targets [17,18]. Increasing the proportion of trips made by walking is one of the fifteen objectives, as recommended by the U.S. Department of Health and Human Service [18].

Social benefit is another outcome of increasing walkability. It is found that walkability is linked to the component of social capital and quality of life in many studies [19]. Social capital is the networks of relationships among people, which can enhance the efficiency of society [20,21]. Studies have also found that neighborhoods with better conditions of walking have higher levels of social capital, as people there are more likely to know each other, and to be more socially connected [22]. In addition, high walkability is always associated with an attractive urban built environment, which can be a driver to restore parts of the cities.

The third is about economic benefit. A highly walkable community could provide a safe and convenient environment for walking and riding, as such, people are likely to spend less on transportation ([www.pedbikeinfo.org](http://www.pedbikeinfo.org)). Health cost could be saved due to sufficient physical exercises [23]. In addition to personal cost, non-motorized travelling mode may reduce public costs for road and parking services, traffic congestion, and environmental damages [23–25]. Besides, walkability could affect housing prices to some extent [26–28]. In the U.S., houses located in more walkable neighborhoods are about \$4000 to \$34,000 more expensive than similar houses in less walkable areas [29]. Further, walkability can promote spending on local businesses to accelerate economic development in communities. Major beneficial activities, such as socializing, shopping, or dining, may appear in pedestrian environment. These activities are directly influenced by the quality of the walking environment [23]. The 2012 Benchmarking Report on Bicycling and Walking in the U.S. [30] found that pedestrian and bicycle infrastructure projects create eight to 12 jobs per \$1 million of spending, compared to just seven jobs created per \$1 million spent on road infrastructure projects [31]. An example is that pedestrians spend more money in New York City over the course of a week than users with any other transportation mode [32].

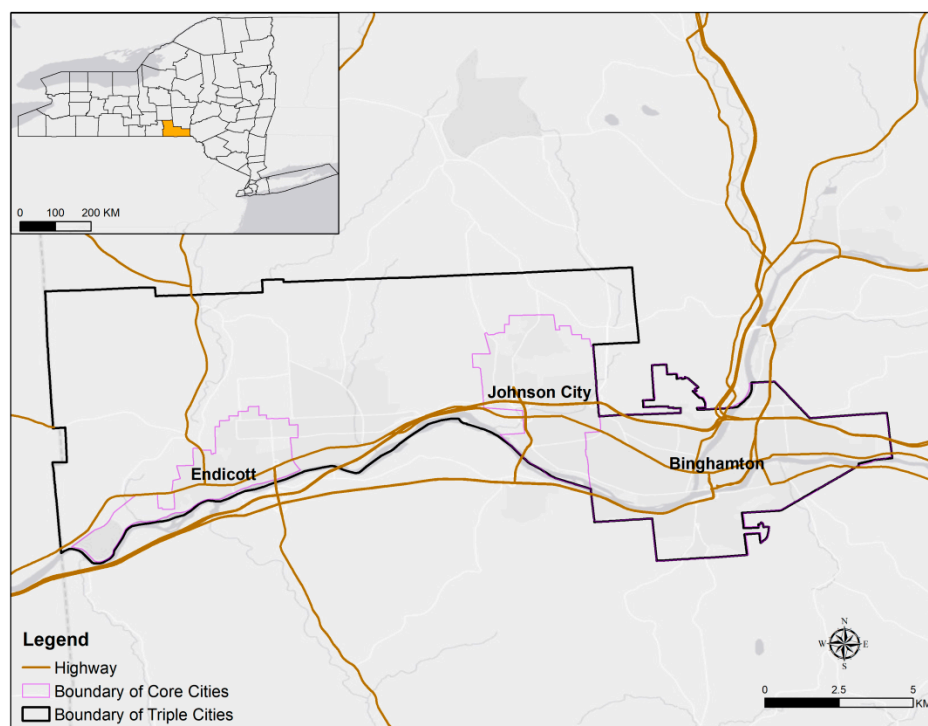
To perform in-depth walkability analysis, one option is to objectively evaluate the different aspects of built environment related to walkability. Three dimensions of walking environment in a community, including density, destinations and design (also known as “3Ds”), were summarized in the literature [33]. Following this fundamental concept, the Neighborhood Environment Walkability Survey was conducted in San Diego, California [6]. Two major aspects of walkability, including connectivity and proximity to destinations, were characterized in [2]. Further, an index was developed at the census block group level to assess the neighborhood walkability by considering density, land-use mix and street connectivity [34]. The Neighborhood Environment Walkability Scale was further improved to measure residents’ perceptions to different environmental characteristics [4,35].

Although existing research effectively reflects walkability to some extent, the metrics in these studies may be further improved in several ways. First, the selection of some variables, for example, the weights used in these indicators, rely heavily on local knowledge, which might be somewhat subjective [34]. Second, some measures used in the literature may not be readily available in other study areas, due to the limited data access as well as regional differences. Third, most of studies in the literature place an emphasis on cities with a large population [33,34,36–47], cities with smaller population and at different development stages are understudied. To address these problems, the

major purpose of this research is to develop a data-driven framework with open datasets, which is able to adaptively refine variables and to quantify walkability in a data-driven manner. It is expected that this data-driven framework will be applicable and useful to any other study area with different features, especially to those post-industrial cities that rarely attract researchers' attentions.

## 2. Study Area and Data

Our study area lies within Broome County, New York, and contains several shrinking cities in the Rust Belt Cities communities historically called the "Triple Cities", and now more broadly as a metropolitan region characterized as the "Greater Binghamton Area" (see Figure 1). The Triple Cities lies within two narrow river valleys (i.e., Susquehanna and Chenango Rivers), with older settlements that tended to be linear until residential expansion pushed development into the hills, sometimes on to relatively steep slopes. Most of the homes and businesses, however, sit on flat terrain, including in the flood plain areas, and in dense or concentrated areas. The entire study area has a large percentage of older structures, including homes built prior to 1950. This includes working class homes built by Endicott-Johnson, Inc. for their employees, and older homes later converted for college students and the poorer populations. The climate typically brings long winters and a late spring, with run-off and freeze and thaw, which can have negative impacts on walkways, sidewalks, and streets. While the sidewalks often provide connectivity between neighborhoods and redevelopment areas, climatic factors can provide challenges. All three cities have aged infrastructure, ranging from buildings to streets and sidewalks, and sewage and water lines.



**Figure 1.** Study area of the Triple Cities (also known as the Greater Binghamton Area), New York.

The old Triple Cities contained the small cities of Binghamton, Johnson City, and Endicott, and still serve as the old urban core of Broome County. These small cities developed proud industrial histories during the industrial era, and some local corporations enjoyed national and global reputations during that period.

Beginning in the late 19th century, producers of candy, cigars, and other products also provided employment, but had less recognition outside of the region. The early 20th Century brought new corporations that received national and international recognition for their industrial quality and offered

large employment. The first, Endicott-Johnson Shoe Corporation (E-J), expanded its employment base to 22,000 workers and had more than 20 factories spread throughout the Triple Cities by the 1940. At its peak it produced 50 million pairs of shoes annually that were consumed globally [48,49]. EJ was a paternal corporation and created a landscape of social and recreational spaces visible today as testimony of their position in the urban region (parks, carousels, and former industrial buildings that are being converted into reusable spaces).

When E-J fell to global cheap labor and competition, the Triple Cities had the good fortune of the arrival of Thomas J. Watson Sr., a friend of George F. Johnson (the owner of E-J), in Endicott in 1906, where he created a worker's time clock. This business evolved into the International Business Machines in the 1920s (i.e., IBM Corporation). After WWII, under the leadership of Thomas J. Watson, Jr., IBM rapidly became a computer industry leader. The second Watson led IBM until the 1960s and the family control of the corporation ceased in the 1970s [50]. IBM had become a high tech corporation that employed nearly 20,000 at its peak. Other tech corporations made the area attractive too, including Link Aviation, Inc. that produced Link's famous Blue-Box Pilot Trainer. That invention found success when the federal government purchased large numbers of Link's machines for training WWII pilots. Binghamton factories boomed and eventually Link Aviation evolved into additional high tech products and was purchased by Singer Corporation.

As a result of the industrial employment provided by these corporations among others, both highly skilled and unskilled work contributed to the growth of Broome County and The Triple Cities until 1970, when Broome County reached its peak of 221,815 (according to U.S. Census Bureau, hereafter). Binghamton had reached its peak population in 1950, 80,674, and continued to fall through 2010 to 47,380. Endicott and Johnson City also reached their population peaks, but in 1970. Since that time, the Triple Cities have evolved into shrinking cities. Manufacturing jobs disappeared; out-migration resulted.

Shrinking cities experience dynamic depopulation trends [51]. Key indicators produce additional shocks that accompany out-migration and continued loss of manufacturing jobs over decades, further reducing chances for revitalization, and include shrinking per capita income, increasing unemployment, declining commercial, and a loss of the tax base. This may be more problematic for small shrinking cities than large ones.

As reported in Table 1, it is clear to observe the trends of depopulation and job losses in the three municipalities, based on the statistics from the 2010 U.S. Census data, which we consider the most accurate when compared to the range of error in the yearly ACS estimates. These trends indicate that the Triple Cities scored high in negative indicators related to the loss of population and employment opportunities. As such, it is very unique to study walking environment under such a social context.

**Table 1.** Negative Trends for Indicators in three cities in the Greater Binghamton Area, 1980–2010.

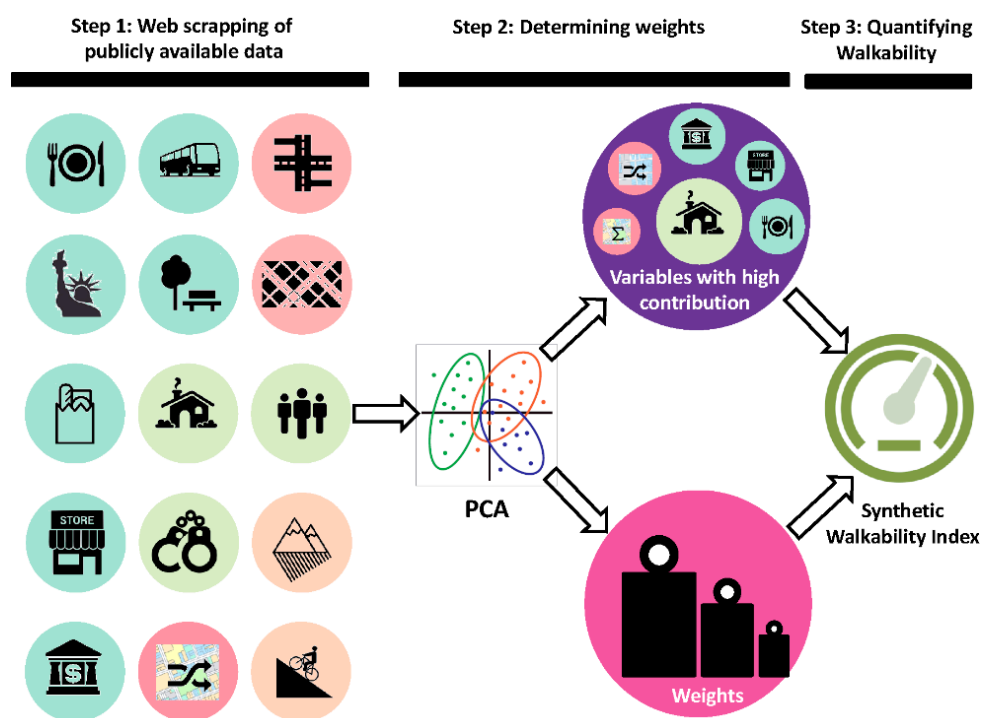
Indicator	Year	Binghamton	Endicott	Johnson City
Percent Unemployed	1980	4.98	5.61	4.4
	1990	4.98	6.42	4.99
	2000	6.47	4.57	6.41
	2010	3.95	6.91	9.2
Percent Below Poverty	1980	30.73	13.2	22.88
	1990	44.37	23.76	17.24
	2000	41.46	23.78	27.98
	2010	49.11	19.24	37.1

In terms of data, multiple open datasets were employed to characterize different dimensions of walkability. The amenity data were used to characterize various types of points of interests (POI) that have either positive or negative impacts on walkability. These POI amenities include fine dining restaurants, fast food restaurants, bars, pubs and taverns, groceries, convenient stores, shopping stores,

health services, banks, auto services, parks, local landmarks, bus stops, and post offices. Addresses of these amenities data were directly extracted from the website of Yellow Pages ([www.yellowpages.com](http://www.yellowpages.com)), and then converted to point data by geocoding. Unlike the POI data used in most studies, which might have been acquired a few years ago, web scrapping of online results can provide the most up-to-date data. More details of web scrapping of these POIs are introduced in the next subsection below. Additionally, the transportation datasets were derived from OpenStreetMap (OSM), where basic geographic information of most countries can be found. They were used to calculate the street density and the street intersections to represent transport connectivity. Digital Elevation Model (DEM) data was used to characterize the elevation and to calculate slope information.

### 3. Methodology

A data-driven framework is proposed in this research. Its architecture is displayed in Figure 2. Three major steps are designed for this framework. The first is to extract open-source data by web scrapping, followed by the calculation of relevant walkability related variables. The second is to perform principal component analysis (PCA) among all variables to identify their respective contribution, and their weights are derived based on the calculated variance. Third, a walkability index of individual location is generated by synthesizing the weights and the contributions of variables. This framework was applied to our study area and described in details in the following subsections.



**Figure 2.** Flowchart of the proposed data-driven framework for walkability quantification.

#### 3.1. Web Scrapping and Measurements Calculation

A variety of variables were derived from and calculated with the use of publicly available open datasets. These variables cover different aspects of urban built environment related to walkability. According to the literature, they include amenities [7,52,53], land use [19,34], transportation infrastructure [54,55], and community environment [56,57] to name a few. The amenities are represented by the density of various POIs, which quantify the potential attractions/distractions of different types of amenities. Such POIs were extracted directly through web scrapping using Python. There are four modules to build a web crawler for scrapping the most up-to-date online POI contents. Web pages were downloaded with query results using keywords (such as “fast food restaurant”), followed by

the parsing location information in each html file. The latitudes and longitudes of target POIs were extracted and converted into shapefile. Web scrapping was repeated to derive all amenity locations (regardless attracting or discouraging walking). We then grouped a collection of amenities with similar functions into a general category. For example, clothing stores, furniture stores and auto stores were grouped as the “shopping stores” class. Hospitals and dental clinics were grouped as “health services”. Overall, seven generalized categories were obtained, including dining services, food markets, shopping stores, daily services, bus stops, local landmarks, and parks. The detailed subcategories of various amenities are listed in Table 2. The amenity density was then calculated as the total count of a type of amenity within a 1-km network-based buffer, following previous studies [19,34].

**Table 2.** Detailed categories of various amenities.

Generalized Categories	Detailed Levels of POIs
Dining services	Full Service Restaurants Fast food Restaurants Bars, pubs and taverns
Food markets	Groceries Supermarkets
Shopping stores	Department stores, furniture stores, clothing stores and discount stores
Daily services	Health services Banks Auto services Post offices
Bus stops	Bus stops
Landmarks	Local landmarks
Parks	Parks

Land use mix was employed to characterize the degree of variation of land uses [34,58]. Shannon diversity index was used to quantify richness and divergence in a group. In this study, this measure was specifically used to quantify the mixture degree of land use types that are closely related to human daily activities. With an emphasis on residential, commercial and office lands following the literature, the Shannon entropy index for land use mix is calculated as follows:

$$M = -\frac{\sum_{i=1}^n p_i \log p_i}{\log n} \quad (1)$$

where  $M$  is the degree of land use mix,  $p_i$  is the proportion of land use type  $i$  within a buffer, and  $n$  is the total number of major land use types ( $n = 3$  in this study). Comparing with just the total count of types of mixed land use within a buffer, this metric can more effectively describe the variation due to different geographical areas but with the same number of mixed land use types [34].

In addition, several types of “density” variables were calculated. These include density of housing, crime, population, intersection, and street. Housing density was calculated as the total count of housing within a 1-km network buffer. We derived the amount of residential housing units directly from the parcel and zoning data. Similarly, population density was calculated as the population total over the residential area. However, the Census data only provide demographic information at the Census tract, block group, and block level, and the boundaries of these scales are inconsistent with that of the buffer zone. Therefore, population totals within the 1-km buffer zone were estimated using the housing unit method [59]. With this method, the total amount of population is estimated as the total counts of housing units multiplied by the average person per household, then adding the population of group quarter. This was done by implementing the remotely sensed data assisted housing unit approach.

More details of small area population estimation are referred to the literature [60,61]. Intersection density is a useful indicator of street connectivity [19]. A community with higher degree of street connectivity would encourage residents to walk or cycle frequently [19,34,54,55]. The intersection density is calculated as the total counts of intersections of the street centerlines within a 1-km network buffer, street density is calculated as the total length of street segments in a buffer with the same size. Highways were not considered when calculating these two indicators.

Finally, two topographical variables were included to examine whether walkability is affected by topographical features, especially due to the fact that the Greater Binghamton Area is located in valley. Two variables, i.e., the average elevation and the average slope within a local buffer, were considered. Elevation was directly obtained from the SRTM DEM dataset, and the average elevation was calculated within the immediate 1-km network buffer. Slope was calculated based on DEM, which was further averaged within each buffer zone.

### 3.2. Synthetic Walkability Index from the Proposed Framework

Based on this framework, a synthetic walkability index was developed to quantify the walking environment in the study area. This index can be calculated at individual location. It is synthesized by integrating a number of variables capturing different aspects of walking environment using principle component analysis (PCA). Prior to conducting PCA, communality of each variable was calculated to remove variables with fewer contributions. Input variables with a communality score less than 0.5 were removed. Components with an eigenvalue greater than 1 were identified and kept as the primary components in the pool. The synthetic walkability index was then derived as a weighted sum of all primary components, by using the explained variance as their respective weights. This index can be formulated as follows:

$$S = \sum_{i=1}^n V_i F_i \quad (2)$$

where  $S$  is the synthetic walkability index,  $n$  is the number of components created;  $F_i$  is the score of component  $i$ ;  $V_i$  is the variance of component  $i$  from PCA, and is used as the weight of a component.

### 3.3. Comparative Analysis

To evaluate the performance of the synthetic walkability index based on the proposed framework, values of Walk Score were extracted from its website ([www.walkscore.com](http://www.walkscore.com)). Walk Score is a walkability metric with a range from 0 to 100 in a limited number of countries. Five categories of amenities are used to obtain Walk Score, including retail (e.g., convenience, drug, grocery, and bookstores), educational (e.g., schools), dining (e.g., restaurants), entertainment (e.g., movie theaters), and recreational (e.g., gyms and parks). It is calculated as a weighted score of Euclidean distances to various nearest amenities, but it remains unknown how the weights are decided. This metric has been used in several applications in different fields [7,52,53]. For comparison purposes, five hundred samples of different addresses in the Triple Cities Area were randomly selected. Values of Walk Score of these random samples were retrieved using the API provided by their website. The synthetic walkability indices of these samples were then compared with the values of Walk Score at the same location. Comparisons between synthetic walkability index and Walk Score were performed for evaluation by using the same random samples.

## 4. Results

### 4.1. Results of PCA

Communality of the input variables was calculated prior to performing PCA (see Table 3). Variables with a threshold less than 0.5 were removed, and the rest were kept for further analysis.

**Table 3.** Variable selection based on communalities.

Variables	Communality	Communality
	(15 Variables)	(11 Variables)
Housing density	0.935	0.955
Population density	0.928	0.951
Land-use mix	0.896	0.905
Dining services	0.884	0.895
Bus stops	0.836	0.878
Crime density	0.803	0.797
Intersection density	0.789	0.878
Food markets	0.762	0.787
Shopping stores	0.748	0.71
Daily services	0.745	0.787
Street density	0.603	0.748
Parks	0.424	/
Local landmarks	0.381	/
Slope	0.238	/
Elevation	0.217	/

Table 4 lists the initial eigenvalues and total variance explained by each component. Four major components were selected as shown as bold fonts. Table 5 displays the rotated loading of each variable to each component. A rule of thumb is that loadings higher than 0.70 are considered excellent, between 0.55 and 0.7 are good, between 0.45 and 0.54 are fair, and lower than 0.45 are poor [50]. The first component has strong positive loadings (all higher than 0.8) on four POI variables: dining service, food markets, shopping stores and daily services. The first component is “amenity attractions”. With high positive loadings on housing density (0.967), population density (0.966), and crime density (0.686) the second component characterizes the “urbanization status”. The third component has strong positive loadings on “transportation connectivity”, with intersection density (0.901), bus stops density (0.880) and street density (0.753). The fourth component has strong positive loadings on land use mix (0.807), which can be interpreted as “land use diversity”. In addition, the correlation coefficient matrix in Figure 3 also supports the grouping of these variables in different components. Variables of amenity attraction have significant positive relationships with each other, with correlation coefficients ranging from 0.484 to 0.799. The two variables of urbanization density, i.e., housing density and population density, highly correlate to each other (with a correlation coefficient of 0.725). The three transportation variables, i.e., street intersections, intersection density and bus stops density, also correlate to each other with correlation coefficients ranging from 0.363 to 0.812. Nevertheless, the relationships between elevation/slope and other variables are not very strong, indicating that elevation might not have a great impact on walkability in this study area.

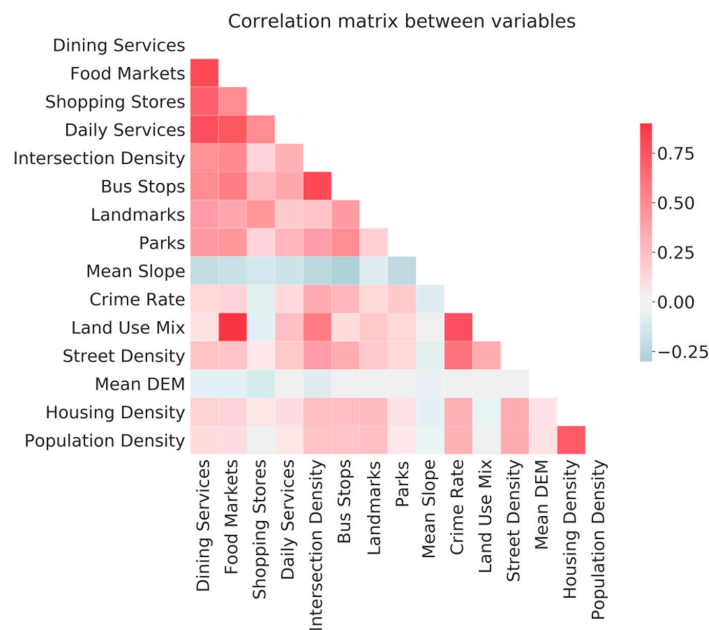


**Table 4.** Eigenvalues and variances (components with an eigenvalue greater than one are bold).

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	4.091	37.187	37.187
2	2.331	21.195	58.382
3	1.403	12.759	71.141
4	1.118	10.16	81.301
5	0.649	5.904	87.205
6	0.447	4.065	91.269
7	0.351	3.191	94.46
8	0.242	2.199	96.659
9	0.2	1.822	98.481
10	0.114	1.034	99.515
11	0.053	0.485	100

**Table 5.** Rotated loading matrix

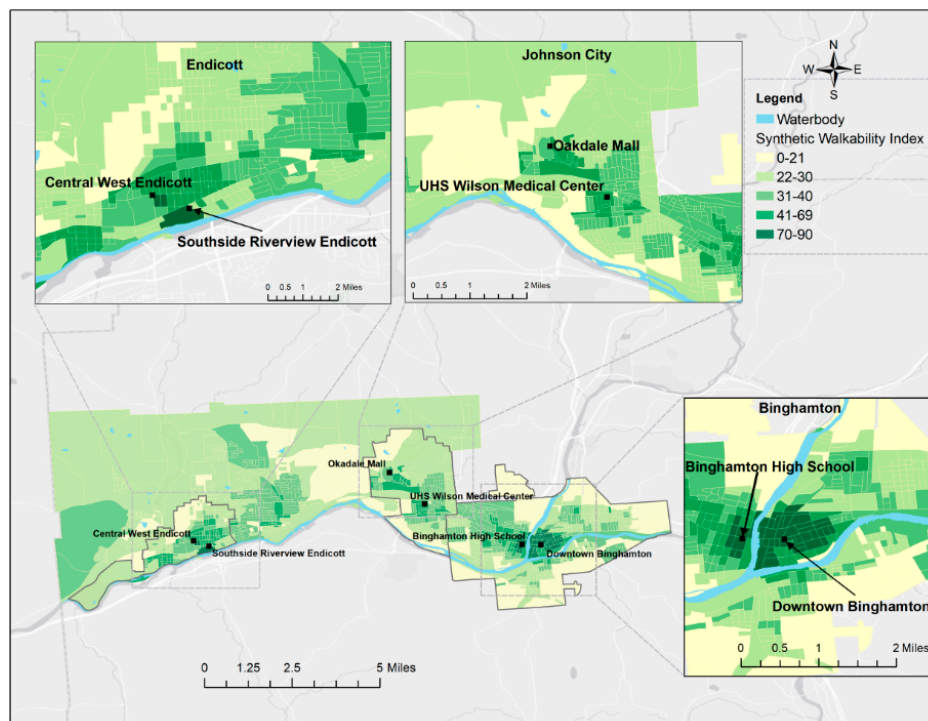
Variables	Components			
	1	2	3	4
Dining service	0.904	0.086	0.262	0.037
Food markets	0.811	0.028	0.359	-0.008
Shopping stores	0.842	-0.015	-0.015	-0.002
Daily services	0.879	0.038	0.134	0.069
Intersection density	0.196	0.07	0.901	0.153
Street density	0.076	0.291	0.753	0.302
Bus stops	0.293	0.115	0.88	0.069
Land use mix	-0.01	-0.173	-0.168	0.807
Crime Density	0.027	0.768	0.274	0.301
Housing density	0.037	0.967	0.084	0.105
Population density	0.041	0.966	0.081	0.097



**Figure 3.** Pearson’s correlation coefficient matrix for variables derived from online open data.

#### 4.2. Results of Comparative Analysis

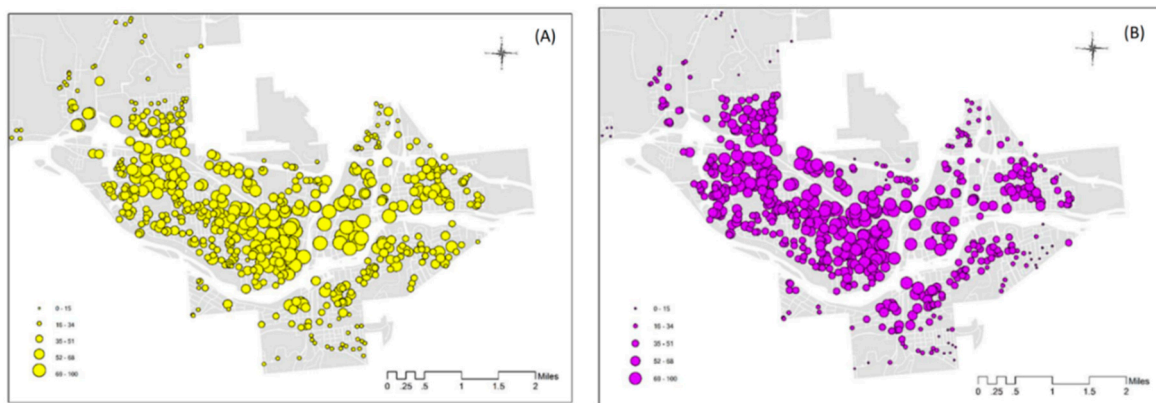
Visual comparison and correlation analysis were used to evaluate the performance of the synthetic walkability index. For illustration purposes, the values of synthetic walkability index of all addresses in each block were averaged. A map showing this new index at the block level was created to display the patterns in the Greater Binghamton Area (see Figure 4). The overall map at the base of Figure 4 shows that most areas of this study area have a relatively low level of walkability (shown as light green). It can be observed that the blocks with a high value of synthetic walkability index were concentrated in the urban center of each of the three cities. As the distances to urban centers increase, the index values substantially decrease. Most suburban areas have a low index value, which indicates that such areas are highly car-dependent. The three detailed inset maps show that the downtown Binghamton, the Central West Endicott and the Southside Riverview Endicott were all highly walkable. In the inset map on the right, the areas in downtown Binghamton tend to be the most walkable area of this region (with the highest values shown as dark green). As the commercial and political center of Broome County, downtown Binghamton is occupied by a large number of restaurants, hotels, shopping stores and government services. In addition, the Greater Binghamton Transportation Center and Binghamton City School District are also located in downtown. As such, it is not surprising that downtown Binghamton is highly walkable. In Johnson City (see the upper right inset in Figure 4), the blocks around the neighboring Oakdale Mall and the UHS Medical Center display a moderate level of walkability. There are also numerous food markets and shopping stores located around these two amenities. Walkability was affected by the highways (as shown in Figure 1) which cut across the Johnson City somewhat reduced the pedestrian connectivity. In downtown Endicott (see the upper left inset in Figure 4), since many small businesses and daily services are located around the Central West Endicott, Southside Riverview Endicott, the synthetic walkability index is also high in these areas along the Susquehanna River.



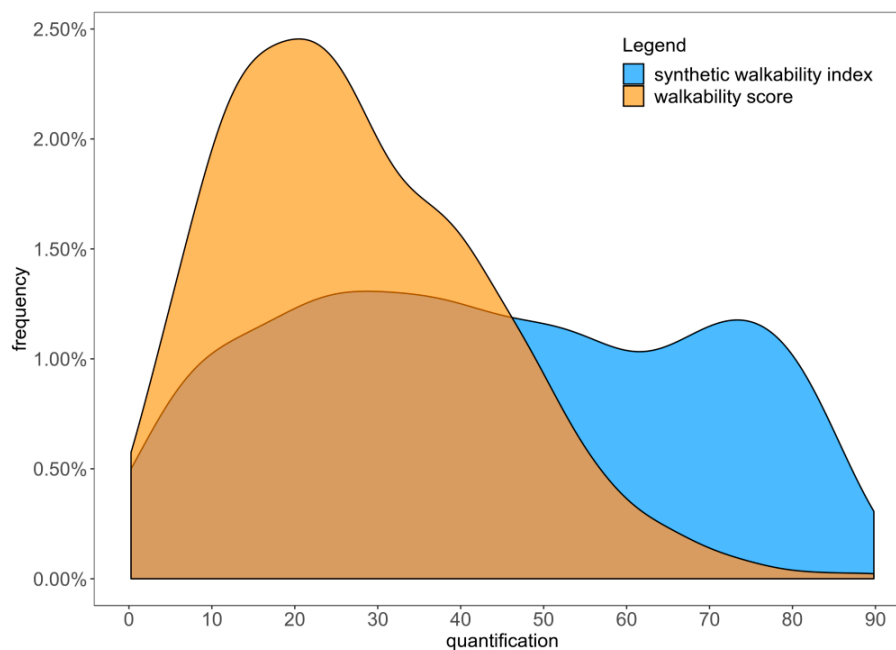
**Figure 4.** Spatial pattern of synthetic walkability index of the Greater Binghamton Area (showing the mean value at the block level, for illustration purposes only).

Figure 5 illustrates the spatial pattern of the 500 random sampled of individual locations (in contrast to the block level) of both the proposed index and Walk Score in Binghamton. Visual

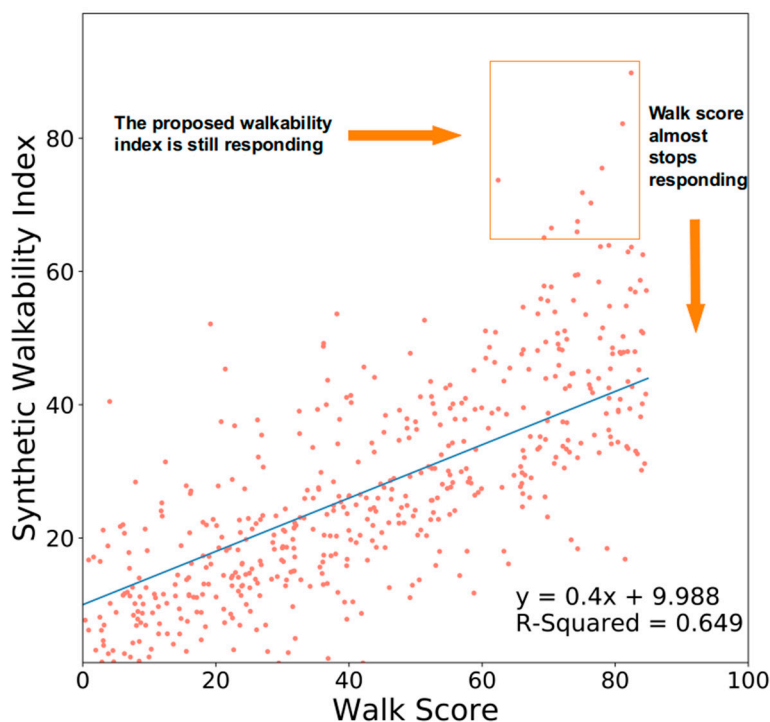
comparison finds that the overall distribution of these two indices is similar: high values in the central (along major roads and in downtown area), and low values in the south. To further examine the relationships between these two indicators, the synthetic walkability index was plotted against Walk Score. In addition to Figure 5, the distribution of this index is also created and shown in Figure 6. Overall, about 25% of blocks are highly walkable with a walkability index greater than 70, while the remaining 75%, of blocks are highly car-dependent, a typical pattern of a shrinking metropolitan area in the Rust Belt in the US. Such a pattern, however, appears to be better reflected by synthetic walkability index (blue) than that of walk score (orange), as the latter one underestimates the amount of highly walkable areas. For further comparison, synthetic walkability index is plotted against Walk Score. Figure 7 shows that most of the scatters cluster along the regression line, with a relatively high correlation of 0.649 ( $p < 0.01$ ). Both the scatterplot and correlation coefficient indicate that there is a significantly positive relationship between the proposed index and Walk Score.



**Figure 5.** Comparison between two indicators: (A) synthetic walkability index; and (B) Walk Score with the same random samples at each individual location in Binghamton.



**Figure 6.** The distributions of synthetic walkability index (blue) and Walk Score (orange) with the same random samples in Binghamton.



**Figure 7.** Scatterplot of synthetic walkability index against Walk Score using 500 random samples at individual location in the study area.

## 5. Discussion

By developing a computational framework and a new walkability metric, the contributions of this research include: (1) The provision of an explicit mechanism for walkability measurement comparable with existing indicator, (2) the data-driven nature for adaptive determination of weights from open data, and (3) the extension of a topical research area in the literature.

### 5.1. Further Comparisons with Existing Indicator

Despite the comparable results and high correlation between the synthetic walkability index and Walk Score, it can be clearly observed in Figure 7 (as highlighted with the orange boxes and arrows) that Walk Score tends to stop responding at locations with relatively high walkability (e.g., it does not go beyond a score of 80), where the synthetic walkability index remains responding. This may suggest a higher sensitivity of the proposed synthetic walkability index than Walk Score in our study area. Also, by examining the scatterplot in Figure 7 together with the distributions in Figure 6, Walk Score seems to slightly underestimate most of the locations in our study area, while the synthetic walkability index, on the contrary, appears to match our local knowledge that three urban cores of this area should be associated with a relatively high walkability value. In New York State, depopulation occurred in a range of cities of the state's regions, including the New York City metropolitan region and Upstate New York communities. The impacts are striking in many ways, including differential reinvestments by both public and private sectors. This realization led to Governor Cuomo's recognition that Upstate communities lagged behind Downstate communities, and a State policy was needed to address the differential. A New York Upstate Revitalization Strategy involved a competition for financial awards exclusively for the Upstate area. The grand prize involved an award of \$500 million for revitalization of a single region. That award was made to the Southern Region of New York, which includes Broome County and the Triple Cities. The Southern Tier Regional Development Council created the regional development project. One of the most important goals involved a strategy for creating three innovation districts (or iDistricts), one in each of the Triple Cities, with the intention of redevelopment of the old

urban core by “Placemaking” and “Downtown Revitalization”. This meant locating in the areas in and near the old central business districts (CBD) and their nearby surrounding neighborhoods that had been hard hit by the depopulation that had suffered from the concomitant negatives expressed by the indicators previously noted in Table 1. The literature has increasingly paid attention to the differential spatial impacts within cities, small and large [62,63]. Some neighborhoods within the city suffered more from the forces associated with depopulation and job losses than others. Certainly, these include neighborhoods in old industrial and near-CBD areas. The Council defined the three iDistricts and the newly defined iDistricts contain such neighborhoods. The revitalization plan focuses on the attracting innovative businesses and an innovative anchor that drives the attraction of new business and employment to the area. Binghamton University provides an anchor for each of the iDistricts and became the driver of redevelopment based on its professional schools. This observation from comparisons in Figures 6 and 7 suggests that the performance of Walk Score may be adequate in cities with high population density, but might need further improvements and adjustments in areas with a small population [52], such as the shrinking metropolitan area in this research. In areas with very different social and economic structures, fewer neighborhood amenities can be derived, the primary inputs to calculate Walk Score, which might lead to a less satisfactory estimate in such areas. Alternatively, the synthetic walkability index does not rely heavily on amenity, but also considers the impacts of other relevant factors, such as population density, crime, and land use mix.

In addition, the proposed framework provides an explicit mechanism for walkability measurement. This is done by scrapping online open data followed by data cleaning and data mining processing steps. Its mechanism is relatively simple and straightforward for calculation, especially nowadays the amenity and socioeconomic records become more and more readily available online. Comparatively, some existing walkability metrics tend to have a black-box mechanism. For example, Walk Score is calculated based on the distance from the target location to the closest surrounding amenities by using a distance decay function [7,52]. The details of its complete algorithm, however, remain implicit to the public. Although an application program interface (API) is provided for data retrieval, it only allows limited daily requests. Accordingly, it is less likely to obtain data for a large area (e.g., at the city level, or an even larger geographic coverage). Moreover, Walk Score only supports four countries of the world at the time when this research was conducted, i.e., the U.S., Canada, Australia, and New Zealand. When walkability quantification is, i.e., in great need in certain cities in countries where there is no immediately available walkability measures (e.g., Walk Score), it is practical for researchers to calculate the synthetic walkability index based on the proposed framework, and to provide such information to local officials in support of urban planning and health practices in these regions.

### 5.2. The Data-Driven Nature of this Framework

Compared with previous studies in the literature, another advantage of this proposed framework is the data driven nature to handle geospatial big data. This framework is designed to use a large amount of relevant variables to cover different dimensions of urban built environment related to walkability. When a variable is not available for walkability quantification in a certain case, this aspect that this variable describes may be still characterized by the major component when numerous variables (and some may be highly correlated with each other) are used as inputs of PCA, as illustrated in this research. In contrast, missing a variable may prevent the calculation when using traditional methods and existing metrics. Admittedly, although PCA is a powerful statistical method, the understanding of what the factor is (or what the factor describes) is left to the researcher.

Moreover, the weight of each component is adaptively determined based on this proposed framework. This is done by using the variance derived from PCA as the weights for synthesizing the index. Comparatively, weights are determined in two ways in existing studies. The first is the usage of equal weights, which, for example, is used for Walk Score [52]. It is based on an assumption of equal importance of different categories of amenities. This may not adequately characterize the reality due to their considerably different impacts on walking. The second limitation is the subjective

assignment of weights, which heavily relies on researchers' local knowledge. Especially when studying different cities with different social structures and economic status, it may be very difficult to decide the appropriate weights in different study areas manually. Therefore, this framework helps adaptively determine weights of variables from online open datasets which characterize the varying impacts of different aspects of walkability.

### 5.3. Extension of Topical Study Areas and Future Studies

This study extends the research topical areas in the literature. A variety of metropolitan areas have been investigated in previous walkability studies. Examples include New York City [36], Atlanta [34,37,38], Los Angeles [39], Seattle [19,40], Portland [41], Indianapolis [42], San Francisco [33,43,44], San Diego [45,54,56], Gainesville [46], Minneapolis–St. Paul [47], and Adelaide, Australia [35]. Unlike existing studies using cities with a large population, our study focuses on a metropolitan area with three typical small shrinking cities in the northeastern U.S. with a smaller population and high obesity prevalence. Population size differences are one issue, and depopulation hits harder on cities with a small population. Due to the relatively low socioeconomic status in many small shrinking cities in the U.S. and other countries, the research outcomes from this study could provide guidance for quantifying walkability in a data-driven manner in these post-industrial cities.

In future studies, we will further improve the proposed framework on the following three aspects. First, we will extend this framework with more geospatial big data by implementing automated extraction of multiple open sources, for example, subjective user generated contents [64], including online review and social media. Second, we will test this framework in different cities of the world where the urban designs and structures substantially vary from the U.S. cities. It is challenging to access and determine the availability and compatibility of open sourced data in differing cities, especially when scrapping online contents written in different languages. Third, in terms of clustering POIs, it is worth trying to use the general categories provided by the websites, such as "Trip Advisor", so as to be easier to polarize the detailed groups of POIs.

## 6. Conclusions

A computational framework for walkability measurement is proposed in this research. Three major steps of this framework include web scrapping of publicly available online data, determining varying weights of variables, and generating a synthetic walkability index. This method was implemented in the Greater Binghamton Area in upstate New York, and the result was compared with the existing walkability metric. The results of this research suggest three major conclusions. First, the proposed data-driven framework provides an explicit mechanism with geospatial big data. Specifically, relevant variables of built environment can be derived from extracted online contents, and their varying weights can be adaptively determined. Second, the synthetic walkability index based on this framework is comparable to the Walk Score, and it even tends to have a slightly higher sensitivity than its counterpart from comparisons. Third, this framework was effectively applied in a U.S. shrinking metropolitan area, which extends the topical area in the walkability literature. It indicates that this approach has the potential to quantify walkability in any city, especially cities with small population where walkability has rarely been quantified and studied. Urban planning strategies, such as land development, street density, and public transit, will indirectly affect urban inhabitants' physical activity and obesity. The proposed framework, as well as the synthetic walkability index, provides a general, explicit, and comprehensive method to capture different aspects of the walking environment in a city. For areas that have rare walkability studies, researchers can calculate the synthetic walkability index based on the proposed framework, to assist urban planners, community leaders, health officials, and policymakers in their practices to improve the walking environment of their communities. Examples include making new re-development plans and policies in order to improve the public health of urban inhabitants. Despite the aforementioned strengths, future studies with this data-driven method are warranted in different cities outside the U.S. for a more comprehensive understanding of the proposed framework.

**Author Contributions:** Conceptualization, Chengbin Deng; Methodology, Chengbin Deng, Xiaoyu Dong and John Frazier; Software, Xiaoyu Dong, Hao Wen, and Huihai Wang; Validation, Xiaoyu Dong, and Hao Wen; Formal Analysis, Xiaoyu Dong, Hao Wen, and Huihai Wang; Resources, Chengbin Deng and John Frazier; Data Curation, Xiaoyu Dong and Hao Wen; Writing—Original Draft Preparation, Chengbin Deng, John Frazier, and Xiaoyu Dong; Writing—Review & Editing, Chengbin Deng, John Frazier, Hung Chak Ho, and Louisa Holmes; Visualization, Huihai Wang, Chengbin Deng, Weiyang Lin and Xiaoyu Dong; Supervision, Chengbin Deng; Funding Acquisition, Chengbin Deng and John Frazier. All authors have read and agreed to the published version of the manuscript.

**Funding:** The APC was funded by the President’s Funding Award for Research at Binghamton University, State University of New York.

**Acknowledgments:** The authors would like to thank the support from our university and local communities.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Moura, F.; Cambra, P.; Gonçalves, A.B. Measuring walkability for distinct pedestrian groups with a participatory assessment method: A case study in Lisbon. *Landsc. Urban Plan.* **2017**, *157*, 282–296. [CrossRef]
2. Frank, L.; Engelke, P.; Schmid, T. *Health and Community Design: The Impact of the Built Environment on Physical Activity*; Island Press: Washington, DC, USA, 2003.
3. Cain, K.L.; Millstein, R.A.; Sallis, J.F.; Conway, T.L.; Gavand, K.A.; Frank, L.D.; Glanz, K. Contribution of streetscape audits to explanation of physical activity in four age groups based on the Microscale Audit of Pedestrian Streetscapes (MAPS). *Soc. Sci. Med.* **2014**, *116*, 82–92. [CrossRef]
4. Cerin, E.; Saelens, B.E.; Sallis, J.F.; Frank, L.D. Neighborhood Environment Walkability Scale: Validity and development of a short form. *Med. Sci. Sports Exerc.* **2006**, *38*, 1682–1691. [CrossRef] [PubMed]
5. Doyle, S.; Kelly-Schwartz, A.; Schlossberg, M.; Stockard, J. Active community environments and health: The relationship of walkable and safe communities to individual health. *J. Am. Plan. Assoc.* **2006**, *72*, 19–31. [CrossRef]
6. Brownson, R.C.; Chang, J.J.; Eyster, A.A.; Ainsworth, B.E.; Kirtland, K.A.; Saelens, B.E.; Sallis, J.F. Measuring the environment for friendliness toward physical activity: A comparison of the reliability of 3 questionnaires. *Am. J. Public Health* **2004**, *94*, 473–483. [CrossRef] [PubMed]
7. Carr, L.J.; Dunsiger, S.I.; Marcus, B.H. Validation of Walk Score for estimating access to walkable amenities. *Br. J. Sports Med.* **2011**, *45*, 1144–1148. [CrossRef] [PubMed]
8. Deanna Westby, M. A health professional’s guide to exercise prescription for people with arthritis: A review of aerobic fitness activities. *Arthritis Care Res.* **2001**, *45*, 501–511. [CrossRef]
9. Vogel, T.; Brechat, P.H.; Leprêtre, P.M.; Kaltenbach, G.; Berthel, M.; Lonsdorfer, J. Health benefits of physical activity in older patients: A review. *Int. J. Clin. Pract.* **2009**, *63*, 303–320. [CrossRef]
10. Warburton, D.E.; Nicol, C.W.; Bredin, S.S. Health benefits of physical activity: The evidence. *Can. Med. Assoc. J.* **2006**, *174*, 801–809. [CrossRef]
11. Murphy, M.H.; Nevill, A.M.; Murtagh, E.M.; Holder, R.L. The effect of walking on fitness, fatness and resting blood pressure: A meta-analysis of randomised, controlled trials. *Prev. Med.* **2007**, *44*, 377–385. [CrossRef]
12. US Department of Health and Human Services. *The Surgeon General’s Call to Action to Prevent and Decrease Overweight and Obesity*; US Department of Health and Human Services: Washington, DC, USA, 2001.
13. Ogden, C.L.; Carroll, M.D.; Fryar, C.D.; Flegal, K.M. *Prevalence of Obesity among Adults and Youth: United States, 2011–2014*; US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics: Hyattsville, MD, USA, 2015; pp. 1–8.
14. Ahima, R.S.; Lazar, M.A. The health risk of obesity—Better metrics imperative. *Science* **2013**, *341*, 856–858. [CrossRef] [PubMed]
15. Finkelstein, E.A.; Trogdon, J.G.; Cohen, J.W.; Dietz, W. Annual medical spending attributable to obesity: Payer- and service-specific estimates. *Health Aff.* **2017**, *28* (Suppl. S1), w822–w831. [CrossRef]
16. World Health Organization (WHO). 2017. Available online: <http://www.who.int/mediacentre/factsheets/fs385/en> (accessed on 8 December 2016).
17. Physical Activity Guidelines Advisory Committee. *Physical Activity Guidelines for Americans*; Department of Health and Human Services: Washington, DC, USA, 2008; pp. 15–34.

18. US Department of Health and Human Services; Office of Disease Prevention and Health Promotion. *Office of Disease Prevention and Health Promotion. Healthy People 2020*; US Department of Health and Human Services: Washington, DC, USA, 2000.
19. Frank, L.D.; Sallis, J.F.; Saelens, B.E.; Leary, L.; Cain, K.; Conway, T.L.; Hess, P.M. The development of a walkability index: Application to the Neighborhood Quality of Life Study. *Br. J. Sports Med.* **2010**, *44*, 924–933. [[CrossRef](#)] [[PubMed](#)]
20. Putnam, R.D.; Leonardi, R.; Nanetti, R.Y. *Making Democracy Work: Civic Traditions in Modern Italy*; Princeton University Press: Princeton, NJ, USA, 1994.
21. Sampson, R.J.; Morenoff, J.D.; Earls, F. Beyond social capital: Spatial dynamics of collective efficacy for children. *Am. Sociol. Rev.* **1999**, *64*, 633–660. [[CrossRef](#)]
22. Rogers, S.H.; Halstead, J.M.; Gardner, K.H.; Carlson, C.H. Examining walkability and social capital as indicators of quality of life at the municipal and neighborhood scales. *Appl. Res. Qual. Life* **2011**, *6*, 201–213. [[CrossRef](#)]
23. Litman, T. Economic value of walkability. *World Transp. Policy Pract.* **2004**, *10*, 5–14. [[CrossRef](#)]
24. Litman, T. Transportation cost and benefit analysis. *Vic. Transp. Policy Inst.* **2009**, *31*, 2–5.
25. Murphy, J.; Delucchi, M. A review of the literature on the social cost of motor vehicle use in the United States. *Inst. Transp. Stud.* **1998**, *1*, 15–42.
26. Boyle, A.; Barrilleaux, C.; Scheller, D. Does Walkability Influence Housing Prices? *Soc. Sci. Q.* **2014**, *95*, 852–867. [[CrossRef](#)]
27. Pivo, G.; Fisher, J.D. The walkability premium in commercial real estate investments. *Real Estate Econ.* **2011**, *39*, 185–219. [[CrossRef](#)]
28. Rauterkus, S.Y.; Miller, N. Residential land values and walkability. *J. Sustain. Real Estate* **2011**, *3*, 23–43.
29. Cortright, J. *Walking the Walk: How Walkability Raises Home Values in U.S. Cities*; CEOs for Cities: Washington, DC, USA, 2009.
30. Swanson, K. *Bicycling and Walking in the United States: 2012 Benchmarking Report*; Alliance for Biking & Walking: Washington, DC, USA, 2012.
31. Garrett-Peltier, H. *Pedestrian and Bicycle Infrastructure: A National Study of Employment Impacts*; Political Economy Research Institute: Amherst, MA, USA, 2011.
32. Transportation Alternatives. *East Village Shopping Survey: A Snapshot of Travel and Spending Patterns of Residents and Visitors in the East Village*; Transportation Alternatives: New York, NY, USA, 2012.
33. Cervero, R.; Kockelman, K. Travel demand and the 3Ds: Density, diversity, and design. *Transp. Res. Part D Transp. Environ.* **1997**, *2*, 199–219. [[CrossRef](#)]
34. Frank, L.D.; Schmid, T.L.; Sallis, J.F.; Chapman, J.; Saelens, B.E. Linking objectively measured physical activity with objectively measured urban form: Findings from SMARTRAQ. *Am. J. Prev. Med.* **2005**, *28*, 117–125. [[CrossRef](#)] [[PubMed](#)]
35. Leslie, E.; Saelens, B.; Frank, L.D.; Owen, N.; Bauman, A.; Coffee, N.; Hugo, G. Residents' perceptions of walkability attributes in objectively different neighbourhoods: A pilot study. *Health Place* **2005**, *11*, 227–236. [[CrossRef](#)]
36. Rundle, A.; Roux, A.V.D.; Freeman, L.M.; Miller, D.; Neckerman, K.M.; Weiss, C.C. The urban built environment and obesity in New York City: A multilevel analysis. *Am. J. Health Promot.* **2007**, *21* (Suppl. S4), 326–334. [[CrossRef](#)]
37. Frank, L.D.; Saelens, B.E.; Powell, K.E.; Chapman, J.E. Stepping towards causation: Do built environments or neighborhood and travel preferences explain physical activity, driving, and obesity? *Soc. Sci. Med.* **2007**, *65*, 1898–1914. [[CrossRef](#)]
38. Frank, L.D.; Kerr, J.; Sallis, J.F.; Miles, R.; Chapman, J. A hierarchy of sociodemographic and environmental correlates of walking and obesity. *Prev. Med.* **2008**, *47*, 172–178. [[CrossRef](#)]
39. Wolch, J.; Jerrett, M.; Reynolds, K.; McConnell, R.; Chang, R.; Dahmann, N.; Berhane, K. Childhood obesity and proximity to urban parks and recreational resources: A longitudinal cohort study. *Health Place* **2011**, *17*, 207–214. [[CrossRef](#)]
40. Frank, L.D.; Sallis, J.F.; Conway, T.L.; Chapman, J.E.; Saelens, B.E.; Bachman, W. Many pathways from land use to health: Associations between neighborhood walkability and active transportation, body mass index, and air quality. *J. Am. Plan. Assoc.* **2006**, *72*, 75–87. [[CrossRef](#)]



41. Greenwald, M.; Boarnet, M. Built environment as determinant of walking behavior: Analyzing nonwork pedestrian travel in Portland, Oregon. *Transp. Res. Rec. J. Transp. Res. Board* **2001**, *1780*, 33–41. [[CrossRef](#)]
42. Lindsey, G.; Han, Y.; Wilson, J.; Yang, J. Neighborhood correlates of urban trail use. *J. Phys. Act. Health* **2006**, *3* (Suppl. S1), S139–S157. [[CrossRef](#)]
43. Kockelman, K. Travel behavior as function of accessibility, land use mixing, and land use balance: Evidence from San Francisco Bay Area. *Transp. Res. Rec. J. Transp. Res. Board* **1997**, *1607*, 116–125. [[CrossRef](#)]
44. Cervero, R.; Duncan, M. Walking, bicycling, and urban landscapes: Evidence from the San Francisco Bay Area. *Am. J. Public Health* **2003**, *93*, 1478–1483. [[CrossRef](#)] [[PubMed](#)]
45. Norman, G.J.; Nutter, S.K.; Ryan, S.; Sallis, J.F.; Calfas, K.J.; Patrick, K. Community design and access to recreational facilities as correlates of adolescent physical activity and body-mass index. *J. Phys. Act. Health* **2006**, *3* (Suppl. S1), S118–S128. [[CrossRef](#)]
46. Handy, S.L.; Boarnet, M.G.; Ewing, R.; Killingsworth, R.E. How the built environment affects physical activity: Views from urban planning. *Am. J. Prev. Med.* **2002**, *23*, 64–73. [[CrossRef](#)]
47. Forsyth, A.; Hearst, M.; Oakes, J.M.; Schmitz, K.H. Design and destinations: Factors influencing walking and total physical activity. *Urban Stud.* **2007**, *45*, 1973–1996. [[CrossRef](#)]
48. Davis, H.B. *The Workers and the Industry*; Shoes: New York, NY, USA, 1940.
49. Zahari, G. *Workers, Managers and Welfare Capitalism: The Shoe Workers and Tanners of Endicott-Johnson, 1890-1950*; University of Illinois Press: Chicago, IL, USA, 1988.
50. Klaf, S.; Legette, K.; Frazier, J.W. *Diversity comes to a small city. The Case of Binghamton NY in Multicultural Geographies. The Changing Racial/Ethnic Patterns of the United States*; Frazier, J.W., Margi, F.M., Eds.; SUNY Press: Albany, NY, USA, 2003.
51. Weaver, R.; Bagchi-Sen, S.; Knight, J.; Frazier, A.E. *Shrinking Cities, Understanding Urban Decline in the United States*; Routledge, Taylor & Francis Group: New York, NY, USA, 2017.
52. Duncan, D.T.; Aldstadt, J.; Whalen, J.; Melly, S.J.; Gortmaker, S.L. Validation of Walk Score for estimating neighborhood walkability: An analysis of four US metropolitan areas. *Int. J. Environ. Res. Public Health* **2011**, *8*, 4160–4179. [[CrossRef](#)]
53. Duncan, D.T.; Aldstadt, J.; Whalen, J.; White, K.; Castro, M.C.; Williams, D.R. Space, race, and poverty: Spatial inequalities in walkable neighborhood amenities? *Demogr. Res.* **2012**, *26*, 409–448. [[CrossRef](#)]
54. Saelens, B.E.; Sallis, J.F.; Black, J.B.; Chen, D. Neighborhood-based differences in physical activity: An environment scale evaluation. *Am. J. Public Health* **2003**, *93*, 1552–1558. [[CrossRef](#)]
55. Saelens, B.E.; Sallis, J.F.; Frank, L.D. Environmental correlates of walking and cycling: Findings from the transportation, urban design, and planning literatures. *Ann. Behav. Med.* **2003**, *25*, 80–91. [[CrossRef](#)]
56. Kligerman, M.; Sallis, J.F.; Ryan, S.; Frank, L.D.; Nader, P.R. Association of neighborhood design and recreation environment variables with physical activity and body mass index in adolescents. *Am. J. Health Promot.* **2007**, *21*, 274–277. [[CrossRef](#)]
57. Hoehner, C.M.; Handy, S.L.; Yan, Y.; Blair, S.N.; Berrigan, D. Association between neighborhood walkability, cardiorespiratory fitness and body-mass index. *Soc. Sci. Med.* **2011**, *73*, 1707–1716. [[CrossRef](#)]
58. Christian, H.E.; Bull, F.C.; Middleton, N.J.; Knuiman, M.W.; Divitini, M.L.; Hooper, P.; Giles-Corti, B. How important is the land use mix measure in understanding walking behaviour? Results from the RESIDE study. *Int. J. Behav. Nutr. Phys. Act.* **2011**, *8*, 55. [[CrossRef](#)]
59. Smith, S.K.; Cody, S. Evaluating the housing unit method: A case study of 1990 population estimates in Florida. *J. Am. Plan. Assoc.* **1994**, *60*, 209–221. [[CrossRef](#)]
60. Deng, C.; Wu, C.; Wang, L. Improving the housing-unit method for small-area population estimation using remote-sensing and GIS information. *Int. J. Remote Sens.* **2010**, *31*, 5673–5688. [[CrossRef](#)]
61. Deng, C.; Wu, C. An integrated geographic and demographic approach. *Ann. Assoc. Am. Geogr.* **2013**, *103*, 1123–1141. [[CrossRef](#)]
62. Franklin, R.S. An examination of the geography of population composition and change in the United States, 2000-2010; insights from geographical indicators and a shift-share analysis. *Popul. Space Place* **2014**, *20*, 18–36. [[CrossRef](#)]

63. Franklin, R.S. The demographic burden of population loss in the U.S. Cities, 2000–2010. *J. Geogr. Syst.* **2019**. [[CrossRef](#)]
64. Hu, Y.; Deng, C.; Zhou, Z. A Semantic and Sentiment Analysis on Online Neighborhood Reviews for Understanding the Perceptions of People toward Their Living Environments. *Ann. Am. Assoc. Geogr.* **2019**, *109*, 1052–1073. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).