

# Exploring the influence of land cover on weight loss awareness

Xin Hong · Xinyue Ye

Published online: 28 August 2017  
© Springer Science+Business Media B.V. 2017

**Abstract** Weight loss is regarded as a commitment for more and more individuals in the US, since obesity is a prevalent public health issue in this country. The geographic environment in which people live and self-awareness of weight status are believed to play an important role bodyweight management. Thus, understanding the association between geographic settings and the awareness of weight status is vital to sustaining or improving quality of life. With the emergence of vast geotagged social media data sources, the analysis of “geographical awareness” of weight status in the context of geographic settings leads to a new research avenue. In order to fully understand the awareness of weight status and the relation with the geographic environment, our analysis was based on weight loss related “tweets” (messages sent on Twitter) and National Land Cover Dataset. We present our findings obtained from modeling: (1) the geographic patterns of weight-loss-related tweets, and (2) how land cover changes impact the weight loss related cyberspace message activities using cross tabulation. The results of analysis are summarized with tables and graphics.

**Keywords** Weight loss · Land cover · Social media · Twitter · Geographic information systems

## Introduction

Growing evidence documents that obesity is prevalent in the US. According to the data from Centers for Disease Control and Prevention (2016), over 31.2% of adults in the US were obese in 2010. The rate of obesity had increased to 33.7% in 2014. Obesity could lead to various health issues, such as blindness, kidney failure, heart attack and stroke (World Health Organization 2016).

It is well documented that unhealthy lifestyles, such as physical inactivity and sedentary behavior, have been implicated as major determinants of obesity epidemic (McGee 2005). However, being physically active or inactive is more than a personal decision. Although obesity could be linked to personal gene or cultural customs, it is believed that the environment in which people live has important role in influencing activities choices (U.S. Department of Health and Human Services [US DHHS] 2001). As highlighted by a body of literature, natural environment, such as green space, is negatively associated with obesity prevalence. Communities that have natural environment facilitated with greater amenities and more recreational opportunities can promote higher levels of physical activity (Michimi and Wimberly 2012; Coombes et al. 2010).

Within the scope of natural environment and physical activity, several research has examined the relationship between green spaces and the participation in physical activities at a neighborhood scale.

---

X. Hong (✉) · X. Ye  
Department of Geography, Kent State University, Kent,  
OH, USA  
e-mail: xhong1@kent.edu

Brown et al. (2009) found that the presence of walkable land uses relates to healthy weight in Salt Lake City, Utah. Coombes et al. (2010) suggested that good access to green spaces may help to encourage population physical activity based on a study in urban Bristol, England. Although there are evidences suggesting that availability and accessibility to green space affect population physical activities and thus impact bodyweight management, limitations and conflicting findings have also emerged. Most of the studies that have examined the correlation between natural environment and bodyweight are limited to a local or neighborhood scale. Research of the same topic may generate conflicting findings from one study area to another study area. For example, Maas et al. (2008) found no associations between green space and participation in physical activities among adults in Netherlands. Also, Hoehner et al. (2005) concluded that there is no evidence suggesting that a 5-min walk distance to green spaces associate with higher physical activity levels.

Part of the reasons for the existing research fail to conclude a consistent association between natural environment and physical activity could be the predominant emphasis on physical measures and the overlook of personal awareness in bodyweight management. Physical measures in the research of the relationship between natural environment and physical activity can be the measurement of distance, walkability, or the amenities of green spaces and the time spent on exercises. Nevertheless, availability, accessibility, and usability of green space do not necessarily imply higher levels of physical activities. Self-awareness on weight status can be an important factor in influencing bodyweight, yet it might be overlooked by some existing literature. Overweight individuals need to be aware that the weight status is harmful to their health before they are motivated to lose weight (Desmond et al. 1986). Fan and Jin (2015) found that adolescents with a self-awareness of being obese have higher weight-loss intentions, but the increased weight-loss intention does not transfer to improved dietary habits or more frequent physical activity. Traditionally, individual-level analyses of weight-loss intentions were studied mostly based on the data collected from questionnaires and interview. In many cases, individual-level analyses have to be limited to regional studies because of the tedious nature of collecting a considerable amount of

representative samples (Fan and Jin 2015; Chang and Christakis 2003). However, personal awareness on weight status cannot be completely understood from generalized regional studies, which is another gap in the existing literature. Therefore, a more efficient and consistent method is needed to quantify geographic patterns of weight status awareness.

This paper presents an innovative approach to procuring individual weight-loss-related posts from social media, or specifically “tweets” to represent weight loss awareness. Tweets contain not only self-reported activities and experiences, but also accurate geo-tagged locations. In addition, social media data is free and the spatial coverage is flexible. Spatial coverage ranges from the spatial movement of a single individual to individuals in the entire world as long as the social media activities are accessible. There are sufficient land cover data derived from remote sensing imageries that are open to the public with free access. This supports the potential of studying spatial relationship between geographic awareness of bodyweight management and environmental variations using geographic information systems (GIS) at lower cost. This paper tends to explore such potential by researching the linkage between weight-loss-related tweets and land cover.

The purpose of this paper is to explore associations between land cover and awareness of weight loss. To the knowledge of the authors, there is no current research investigating personal recognition of weight status using social media data. Weight loss awareness is considered a new indicator for public health in this paper. The approach conducted in this paper is considered an innovative and integrated application of big data and GIS in public health studies. Findings of this paper are expected to provide implications to inform health scientists the importance of adopting social media to investigate self-evaluation on personal wellness and the importance of land use to health.

### **Measuring natural environments and bodyweight in pre-defined boundary**

Even though genetic, scultural, and daily diet habits explain bodyweight, the natural environment is often considered a determining factor in bodyweight (US DHHS 2001). In order to probe the association between natural environment and bodyweight, the

majority of studies approach the research question from a rather predefined regional or community geographic boundary. Physical inactivity is found to significantly aggravate obesity, and a majority of research attempts to seek how natural environment might affect physical activities (Coombes et al. 2010; Michimi and Wimberly 2012; Wolch et al. 2011). Natural environment in the context of exercise levels is conventionally represented by the accessibility and the quality of natural amenities, such as walkability/accessibility from homes to recreation facilities and the green cover of the facilities (Thornton et al. 2011). Physical activity and bodyweight are commonly estimated by data collected from questionnaires or secondary data from governmental agencies. Wolch et al. (2011) assessed the influence of proximity to recreational resources on childhood obesity on 12 communities in Southern California from written questionnaire databases. Brown et al. (2009) investigated the linkage between mixed land use and walkability and obesity using data acquired from Census 2000 and driver licenses in Salt Lake City. Samples collected from individuals, especially the data used to indicate bodyweight status and physical activities, are suitable for evaluating health issues in great detail, though they are difficult to acquire in a larger geographic extent. Consequently, spatial patterns of weight-related issues from a large geographic perspective are challenging to research due to the fact that data acquisition is a tedious process in individual-level analysis.

### Weight loss awareness

As discussed in the previous section, studies of bodyweight and its association with natural environment are restricted to a predefined regional boundary and lack of the depiction that beyond a regional scale. Another flaw of current literature is the neglect of awareness or self-recognition. It is well documented that higher levels of physical activity/exercise are correlated with clinically-significant weight reduction (Butler et al. 2004; Drewnowski and Rolls 2012). Nevertheless, studies of the environmental influences on discouraging or encouraging physical activities are mainly based on physical measures. Even there are few studies attempting to extend the study area to national scale, bodyweight status is always drawn

from physical measurement. For example, Ghimire et al.'s (2015) investigation on green space and adult obesity prevalence in the US. Obesity prevalence in that study was represented by the integration of green space percentage, public open space accessibility, unhealthy eating habits, etc. The same issue for regional studies (including the studies aforementioned in the previous section) on the measure of bodyweight and green space accessibility. Bodyweight status is traditionally present using body mass index (BMI) and height (Frank et al. 2006). However, BMI may not be the criteria to estimate bodyweight, because BMI fails to account for body composition. For instance, high BMI values might reflect only greater muscle mass rather than obesity (Michimi and Wimberly 2012).

The over-emphasis of physical measures on bodyweight and natural environment tend to disregard the importance of intrinsic awareness of weight status in promoting motivation for weight control. It is a common sense that physical activity/exercise is positively associated with successful long-term weight control. It was recommended that additional 60–90 min per day of moderate intensity physical activity to a dietary intervention can substantially promote successful long-term weight control (Donnelly et al. 2009). Surprisingly, in the context of long-term weight management, more than 70% of US adults fail to meet this recommendation and about 80% of individuals seeking weight loss are unable to successfully integrate physical activity into their lifestyles (Spiegel and Alving 2005; Wing and Hill 2001). The issue remains in how individuals can optimally facilitate the adoption of a physically active lifestyle. Silva et al. (2010) argue that motivational dynamics of exercise and weight loss initiation and persistence should be given special focus, though there are other determinants could intervene healthy lifestyles. Many studies have shown that individuals with awareness of overweight or obesity tend to have stronger motivation for losing weight and are more autonomous attending weight control programs (such as physical exercise and healthy diets), lose more weight during the program, and maintain weight loss status better in a long term (Williams et al. 1996; Silva et al. 2010). Therefore, understanding how natural environment influence awareness of weight loss is a prerequisite to better promote weight loss motivation.

A growing trend in applied geography has focused on the use of user-generated web contents in a bottom-

up approach from mobile sensors carried humans via social media (Arribas-Bel 2014). With the emergence of “Web 2.0”, social media such as Facebook, Twitter, and Flickr, where users can generate and modify online contents, have become new instruments for collecting health information and keep track of individual awareness of disease prevalence (Robillard et al. 2013). With 342 million active members and average 58 million tweets per day by 2016, Twitter has kept attracting new users with the increasing popularity of smartphones (Twitter 2016). This new kind of data has been adopted to explore spatial and temporal dynamics in medical/health geography with the integration of GIS, for example food environment (Chen and Yang 2014), dementia (Robillard et al. 2013), and influenza (Paul and Dredza 2011). We believe that the applications of geotagged social media data can also extend to its correspondence with natural environment. In this paper, we propose to harness weight-loss-related tweets and conventional land cover data, which both tightly tether to geospatial coordinates, to explore the correspondence between weight loss awareness and land cover types using GIS.

The fusion of Twitter messages acquired from “human sensor” and land cover information as collected from traditionally remotely sensors in this paper introduces a novel manner of researching health geography. The originality of this paper is embodied in: (a) the introduction of geographic awareness of weight loss, as reflected by voluntarily individual tweets, supports bodyweight themed analyses; (b) the

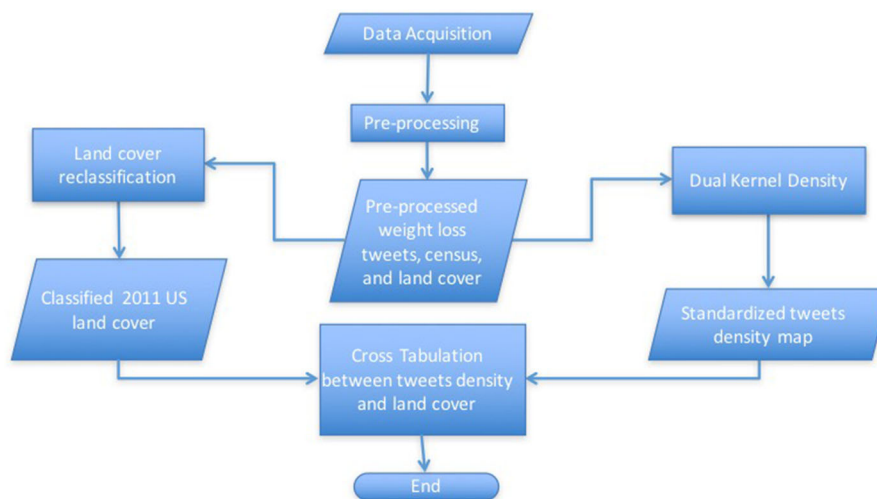
exploration of spatial dynamics between weight loss awareness and land cover variations opens a platform for the marriage between newborn big data and traditional remote sensing data; (c) the demonstration of using tweets and land cover data to facilitate health studies for a large study area in an economic and rather simple way.

## Data and method

The objective of this paper is to explore the influence of land cover variations on perceived weight loss in the US. To address this, a three-phase methodology was conducted by: (a) mapping the spatial patterns of weight-loss-related tweets using Dual Kernel Density Estimation (KDE); (b) reclassifying the 2011 National Land Cover Database; and (c) cross-tabulating the relationship between weight loss tweets and land cover. The workflow of this paper is displayed in Fig. 1.

### Collecting weight loss conversations on Twitter

Twitter Search Application Programming Interface (API) (<https://search.twitter.com/>) was used to extract weight-loss-related tweets. The extraction was started with searching tweets posted during January 7 to February 14, 2016 (5 weeks) containing defined terms: 1. “losing weight”, 2. “losing weight” and “lost pound”, 3. “losing weight” and “workout diet”,



**Fig. 1** Flow chart of the methodology

4. “losing weight” and “workout fitness”, 5. “losing weight” and “workout fit”, 6. “losing weight” and “dieting”, 7. “losing weight” and “gym fit”, 8. “losing weight” and “gym fitness”, and 9. “losing weight” and “workout dieting”. These search terms were chosen based upon the domain weight-loss-related literature, which the majority of them are within the scope of “workout”, “die”, “fitness”, “gym”, etc. (Butler et al. 2004; Maas et al. 2008; Silva et al. 2010; Wharton et al. 2008).

The next step was to “clean up” the original tweets, including deleting the re-tweets and the tweets geotagged at the geographic center of the US. A total of 30,081 tweets remained after filtering. Based on the inherent XY coordinates in the tweets, locations of the tweets as displayed as point features within the US 48 contiguous states (plus Washington, D.C.) (Fig. 2).

National Land Cover Databases 2011 (NLCD 2011) was retrieved from the Multi-Resolution Land Characteristics Consortium (MRLC <https://www.mrlc.gov/nlcd2011.php>). NLCD, derived land cover from Landsat scenes, explicitly depicts the US land cover information at a spatial resolution of 30-m. In total, 20 land cover types were classified in the NLCD and the classes focus on the description of present natural materials and intensity of human modifications on the Earth surface (MRLC 2016). The decent spatial resolution and detailed land cover information of

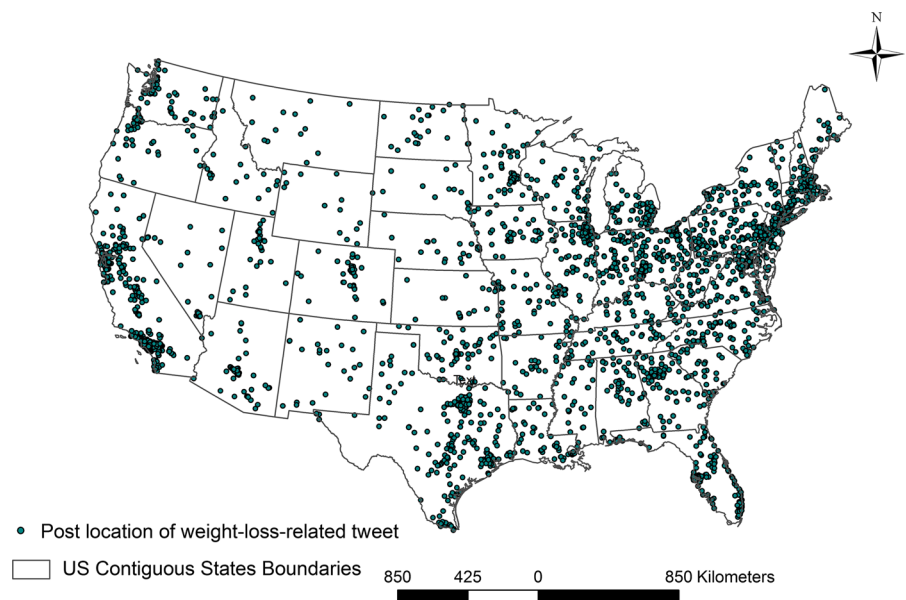
NLCD facilitate the coupling of tweets and remote sensing data in national-scale research of associations between human perceived health conditions and geographical variations.

Description of each land cover can be found on the NLCD website ([https://www.mrlc.gov/nlcd11\\_leg.php](https://www.mrlc.gov/nlcd11_leg.php)). The 20 land cover types include both the first and second hierarchical categories. In order to avoid redundant information from the original land covers, the land cover types employed in this study were grouped up to the 8 first hierarchy classes: 1. water, 2. developed, 3. barren, 4. forest, 5. shrubland, 6. herbaceous, 7. planted/cultivated, and 8. wetlands.

#### Dual kernel density

There is a variety of spatial tools to analyze spatial patterns of point features, such as the points of tweets shown in Fig. 2. Kernel density estimation (KDE) was chosen in this study to map the changing geographies of tweets. KDE is an interpolation technique that generalize the distribution pattern of point incidents to an entire area and illustrates the density estimates (Fischer et al. 2001). Compared to statistical hotspot and clustering techniques, the main advantage of KDE lies in determining the occurrence of likelihood. The occurrence of likelihood of tweets can be defined as an area of clustering in which there is an increased

**Fig. 2** Geotagged weight-loss-related tweets within the US 48 contiguous states (plus Washington, D.C.) posted during January 7 and February 14, 2016 (after deleting re-tweets and the tweets located in the US geographic center)



Xin Hong  
11-29-2016



likelihood for a tweet to occur based on spatial dependency (Anderson 2009). There are two types of KDE: a single variable density estimate that one variable is applied; and a dual density estimate that two variables are applied. In the latter case, a kernel density estimate is firstly applied to both variables respectively, and then these two estimates are associated with each other by a simple algebraic operation, such as ratio, difference, or sum (Jansenberger and Stauffer-Sterinnocher 2004; Levine 2013).

The density of tweets can be influenced by population density. Therefore, to reduce the effect from population, population at census tract level was introduced to standardize the tweet distribution. An example of using population to standardize social media data can be found in Wang et al.'s (2016) spatial analysis of Twitter conversing for the topic of wildfire hazards. In the context of this study, a Dual KDE, which calculates the ratio densities of tweets and population densities, was harnessed to generate a tweet density map. The formula of Dual KDE is:

$$\text{Dual KDE} = \frac{\text{KDE of tweets}}{\text{KDE of population at census tract level}} \quad (1)$$

To be specific, prior to operating the ratio of densities, location of tweets (Fig. 2) and location of

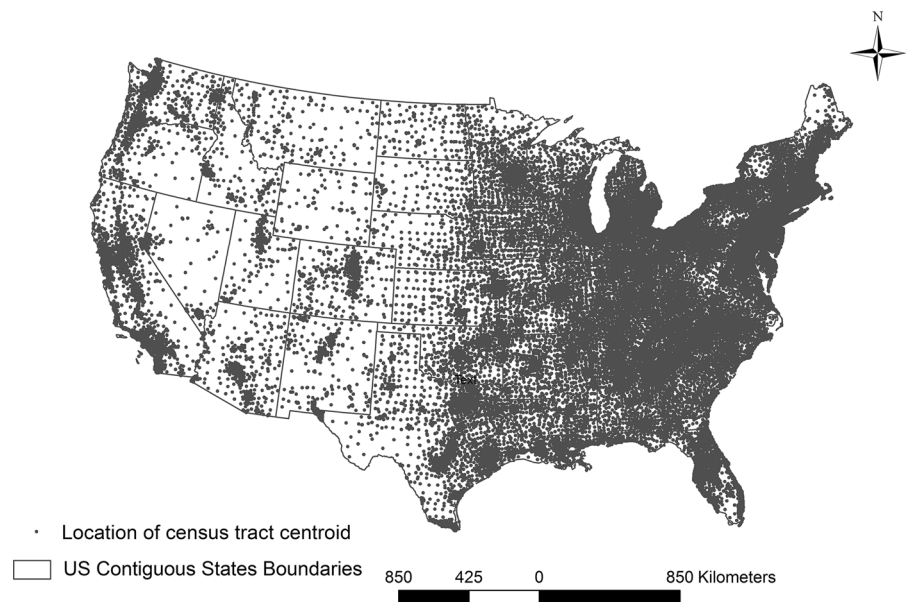
2014 census tract population centroids (Fig. 3), were applied to interpolate the KDE surfaces, respectively. The Dual KDE calculation was implemented in CrimeStat software (Levine 2015) for the ratio operation. Considering the large spatial coverage of the study area and limited performance ability of the software, 10 km was chosen as the spatial resolution for the output Dual KDE raster.

Cross tabulation between tweet density and land cover

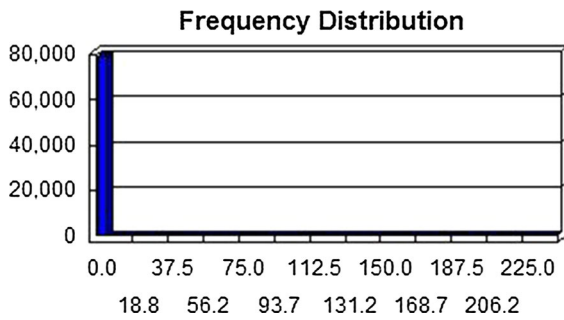
Cross tabulation, a pseudo-arithmetic manipulation logic, was employed to guide the association between tweet density and land cover. Cross tabulation calculates the interrelations between two categorical layers through a two-dimensional contingency table that records the frequency of intersection between their attribute values in a cross-wide comparison (Lein 2012, p. 308). *Tabulate Intersection* tool in ArcGIS was used to perform the cross tabulation between tweet density and land cover. To be specific, *Tabulate Intersection* was harnessed to compute the area and count of the intersecting features between the tweet density and the land cover vectors.

As a preparation for the cross tabulation, the Dual KDE raster was converted to a polygon vector. According to the Dual KDE histogram (Fig. 4), the

**Fig. 3** Points of centroid XY coordinates of 2014 US population at census tract. Source U.S. Census Bureau <https://www.census.gov/>



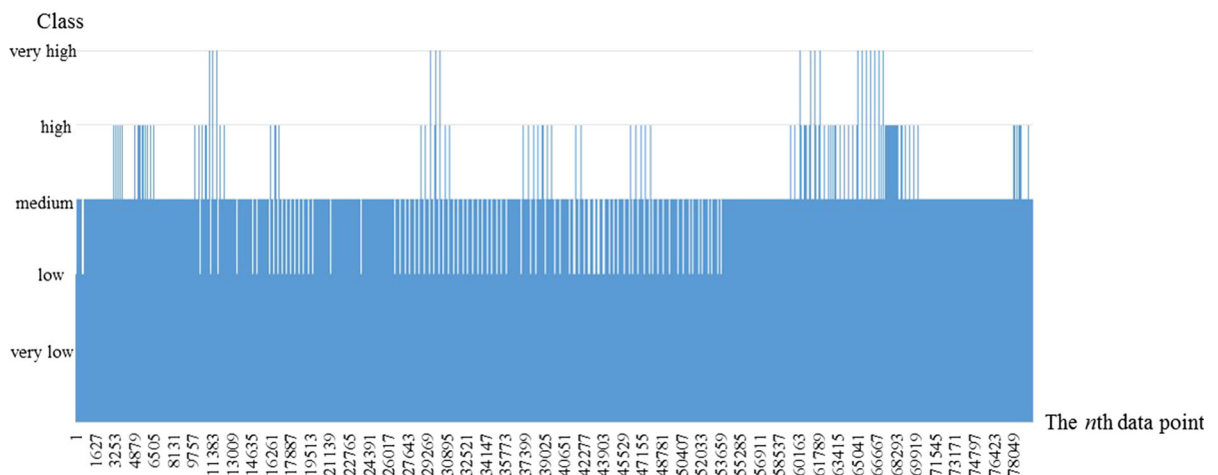
Xin Hong  
11-29-2016



**Fig. 4** Histogram of weight-loss-related tweets based on Dual KDE values

data is heavily skewed by the preponderant values between 0.01 and 0.60. Through practicing different data classification schemes, Natural Breaks (Jenks), quantiles, etc., geometrical interval is better in creating a more even counts of values in each class range and more constant changes between intervals. Geometrical interval creates class breaks based on geometrical series and minimizes the square sum of element per class for data which often do not have normal distribution (ArcGIS 9.2 Desktop Help 2007).

Five tweet density classes based geometrical intervals were divided, their class names and corresponding values are: very low (0–0.058), low (0.059–0.482), medium (0.483–3.837), high (3.838–30.386), and very high (30.387–240.52). Figure 5 illustrates the distribution of data points in tweet density classes. For example, the 8131st data point is categorized in the Medium class and the 11,383rd data point falls in the



**Fig. 5** Distribution of tweet density data classified by geometrical interval scheme

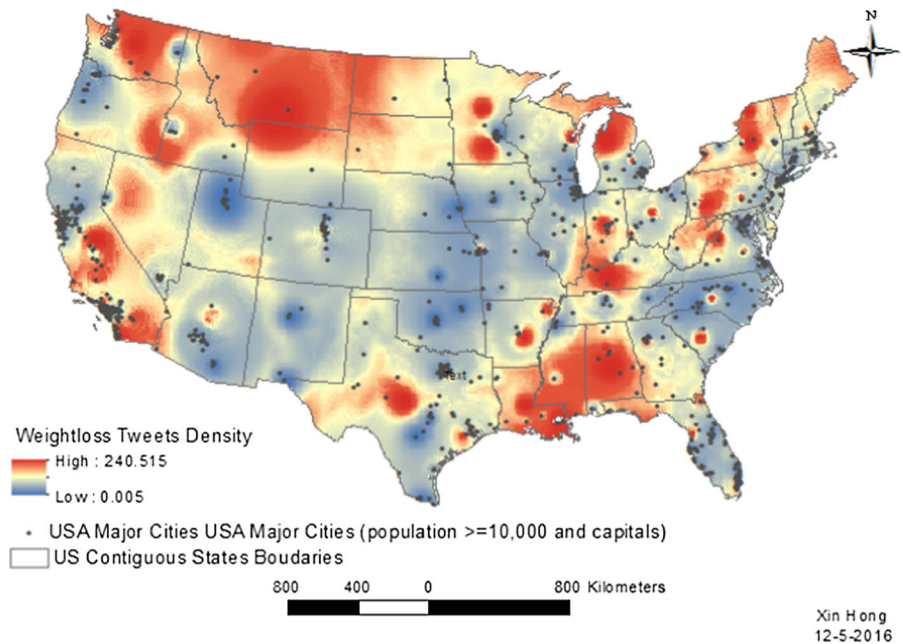
Very High class. By comparing the histogram of the tweet density values (Fig. 4) and the distribution of tweet density classes classified by geometrical interval, we can discern that geometrical interval accommodates the data values with a heavily-skewed histogram to a more even distribution.

### Results

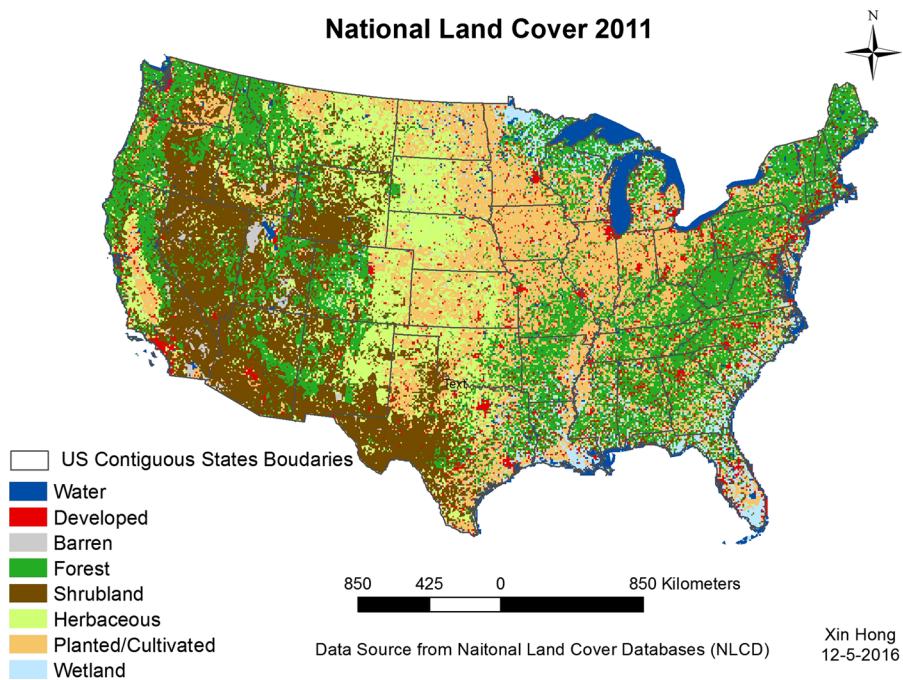
The delineation of weight loss awareness

After removing the effect of population, Dual Kernel Density Estimation (KDE) representing the density of weight loss awareness with US major cities and US state boundaries overlaid is shown in Fig. 6. As seen from the spatial distribution, the popularity of discussing weight-loss-related topics does not necessarily associate with urban and rural locations, in spite of the facts that rural areas have higher obesity prevalence than urban areas in the US (Liu et al. 2008; Michimi and Wimberly 2012). By visually corresponding the Dual KDE map (Fig. 6) with the reclassified national land cover (Fig. 7), higher density of conversing weight loss spread in a similar pattern as forest. Higher concentrations of tweets were mainly located on northwest US (Montana, Washington, and Idaho), south California; gulf-coastal states (Louisiana, Alabama, Georgia, and Central Texas), Michigan, and Appalachian regions. Interestingly, these areas also have higher concentration of forest.

**Fig. 6** Dual KDE map. The ratio of weight-loss tweets standardized by census tract population



**Fig. 7** Reclassified 2011 National Land Cover



Associations between weight loss conversation and land cover

A closer inspection of the Dual KDE (Fig. 6) and the national land cover (Fig. 7) shows higher density of tweets around national forests and parks, including the

Glacier region Montana, the ones which lie in the border between Montana and Wyoming (Yellowstone National Park, Custer Gallatin National Forest, and Bighorn National Forest), Mt. Baker-Snoqualmie National Forest in northwest Washington, Umatilla/Malheur National Forest in Oregon, Nez Perce-



Clearwater/Payette National Forest in Idaho, the forest areas in north Michigan, the hilly regions in southwest Kentucky, De Soto National Forest in south Mississippi, Talladega National Forest in Alabama, the national forests in the outskirts of Los Angeles and San Diego, and the Appalachian mountain regions. Interestingly, those areas are also characterized by tourist and recreational attractions. The tweets were posted during January and February, which is the snowfall season in north United States. Winter recreation, such as winter camping, skiing, and snowmobiling, have brought a considerable number of visitors every year in the aforementioned national forests/parks in north states (US Forest Service 2015). The national forests/parks in south US, such as De Soto National Forest in Mississippi, support year-round recreation opportunities for hikers, campers, canoeists, etc. (US Forest Service n.d.). We assume that higher frequency of conversations related to weight loss are likely aroused by the recreational activities experienced in the forests/parks. Even apart from the activities that the individuals were experiencing, it can be certain that, based on the geo-location of the tweets, individuals in national forests/parks tend to converse more on weight loss.

The precise coordinates of tweets from geo-tagged performance in their mobile devices provide the

feasibility to correspond land cover location to the weight-loss-related tweeting conversation. By cross-tabulating the Dual KDE of weight loss tweets and land cover, a new representation depicting the association between weight loss awareness and land cover is generated. Figure 8 summarizes the percentages of four KDE weight loss tweets categories (very low, low, medium, high, and very high) in each land cover (water, developed, barren, forest, shrubland, herbaceous, planted/cultivated, and wetlands). In general, the intensity of weight loss tweets was in direct proportion to green space (forest, shrubland, and herbaceous). On one hand, higher intensity of weight loss tweets was more likely to occur in green covers. On the other hand, weight loss tweets tended to have an inverse relationship with developed and planted/cultivated cover, respectively, even though the percentage of low intensity tweets was greater than very low intensity tweets presenting in planted/cultivated cover. Approximately 35% of the very high tweet density and 31% of the high tweet density classes overlaid with forest, respectively. Around 18% of high tweet density and 25% of very high tweet density were located in herbaceous. In terms of very low tweet density, a fairly large amount (around 28%) of it overlaid with developed land cover.

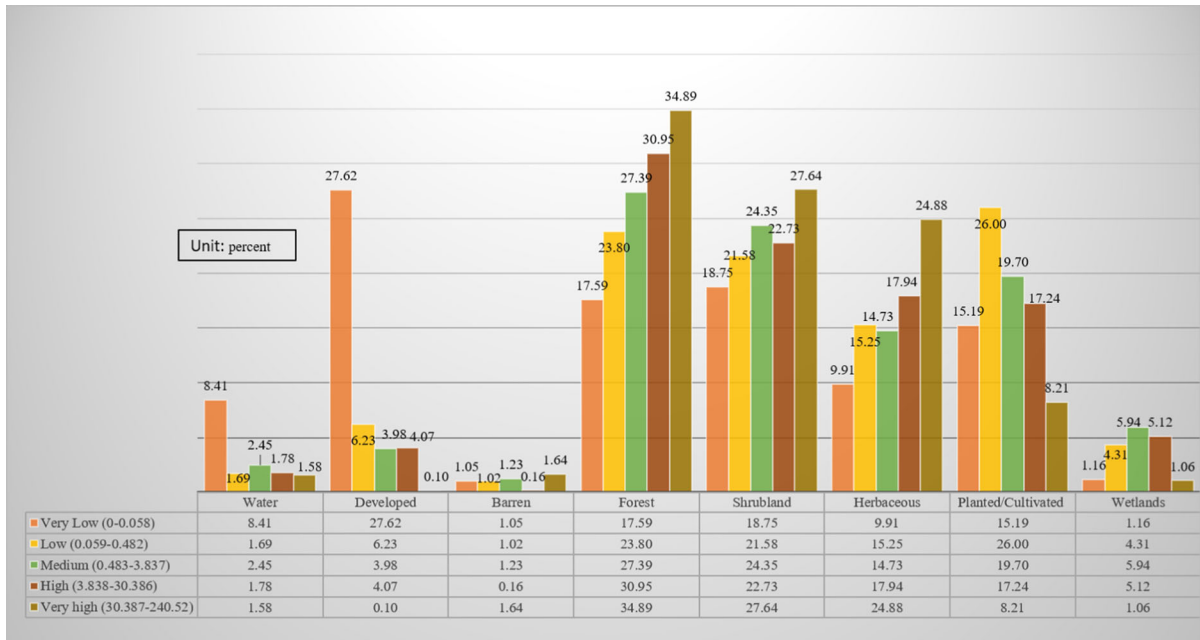


Fig. 8 The percentage Dual KDE of tweets by land cover

**Table 1** The count of weight-loss-related tweets in each Dual KDE class by land cover

Tweet density class	Water	Developed	Barren	Forest	Shrubland	Herbaceous	Planted/ cultivated	Wetlands	Sum by density class
Very low	55	181	7	115	123	65	100	8	654
Low	125	460	75	1750	1587	1122	1912	318	7349
Medium	362	584	183	3999	3555	2153	2877	872	14,585
High	78	177	7	1349	990	782	751	223	4357
Very high	50	3	51	1094	867	780	258	33	3136
Sum by land cover	670	1405	323	8307	7122	4902	5898	1454	30,081

As mentioned earlier in the *Data and Method* section, 30,081 tweets were chosen for the study. Regardless of the limited sample size, the count of tweets falling within each Dual KDE class and the corresponding land cover category is summarized in Table 1. Among all the 30,081 tweets, 8307 tweets were geotagged in forest, which accounts for over 27% of the tweets. Following by forest, shrubland, planted/cultivated, and herbaceous occupied 7122, 5898, and 4902 of the tweet geotagged locations, respectively, which also considerably hold greater number of tweets than other land cover (water, developed, barren, and wetlands).

## Discussion and conclusion

In this study, we focus on using tweet data to map the spatial differences in the level of weight loss awareness and exploring the geographic correspondence between weight loss awareness and land cover. The combined use of topic extraction from big data source, remote sensing, and Geographic Information Systems (GIS) demonstrate the potentials of researching the association between health improvement and natural environment.

Findings of this study have several implications in public health and natural resource management. As indicated by the results, there are obvious differences in the level of geographical awareness of Twitter users between green spaces and non-green spaces. Higher levels of weight loss awareness correspond with green spaces, as forest and shrubland had considerably greater ratio of weight loss tweets compared to other land cover (Fig. 8). In addition, approximately 30% (8307 out of 30,081) of the weight loss tweets were

post in forest, and about 24% (7122 out of 30,081) were in shrubland. Conversely, developed, water, and barren land cover significantly associated with lower weight-loss-related tweet density. It appears that land cover, as one of the geographical variations, does have influence on individual weight loss conversing in Twitter. People tend to be more aware of their weight status and losing weight when they are in green space, especially areas with long-standing reputation in recreational activities/sports, such as Yellowstone National Park crossing Montana and Wyoming and De Soto National Forest in south Mississippi. In contrast, environments in developed and cultivated/planted areas seem to discourage the weight loss awareness among Twitter users. It suggests that being in a green space environment would inspire stronger craving for weight loss conversation in cyberspace, regardless of the activities or practices related to weight loss. Hence, preserving and improving accessibility of green space, especially forest, could promote weight loss awareness.

Existing research has focused intensively on how to facilitate healthy lifestyle in order to reduce obesity prevalence (Thornton et al. 2011; Ghosh and Guha 2013), while rare attention has been given to how to promote individual awareness on weight management prior to engaging weight loss practices. Public agencies may see a health benefit in promoting weight loss awareness by increasing availability and accessibility to public green space, such as national and state parks. Traditional approaches to controlling overweight and obesity in public health practices should be extended to more comprehensive approaches that acknowledge the role of weight status awareness. Introducing more outdoor recreational opportunities in the natural environments, such as improving the amenities for

physical activities in state/national parks, tends to increase weight loss awareness. Moreover, it is well-known that outdoor recreational activities also help to maintain active living. This paper opens a venue for researching the dynamics between geographic awareness of weight status and natural environment. It explores the potentials of big data with conventional land cover data combined in public health studies using GIS. Outreach researching topics extended from this paper could be the interactive dynamics between weight loss awareness and physical activities in green spaces for improving weight management practices. For instance, improved recreational opportunities in green spaces may promote both physical activities and bodyweight awareness; and the increased bodyweight awareness would interactively encourage higher levels of physical activities.

To the knowledge of the authors, this paper is the initial attempt of adopting online crowd communication to represent the levels of geographic awareness of weight loss via self-reported cyber activities. This study underscores the potential of positioning social media as a data source in health science study on the topics related to personal wellness. In addition, this study highlights the convergence between conventional and newly emerging data sources with the attempt of coupling remote sensing data with Twitter data. Hopefully, this research will help to evaluate land value in public health from individual conversations in the dense and complex informational cyberspace.

However, several limitations should be taken into account in the future research. First, the demographic composition of Twitter users was not considered in this study. Twitter users might be dominated by young population. Conclusions drawn from this research may not be able to represent the entire population. Provided that demographic information of Twitter users was acquired, more comprehensive investigations on the impacts of socioeconomic variations on weight status and obesity prevalence could be achieved. Second, self-reported weight loss tweets might only represent weight loss awareness. We are not able to determine whether weight loss tweets were a valid measure of the actual weight loss experience, although the study finds there is an association between weight loss awareness and green space. Third, due to limited capability of software, Dual KDE and land cover must be converted to coarser resolution (10 km) prior to implementing

cross-tabulation for the entire US. Conclusions from this study should also account for the compensation for variations in local areas.

**Funding** This study was funded by the National Science Foundation under Grant No. 1416509.

#### Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Human and animal rights** This paper does not contain any studies with human participants performed by any of the authors.

## References

- Anderson, T. K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident Analysis and Prevention*, 41(3), 359–364.
- ArcGIS Desktop Help 9.2. (2007). *Geometrical interval*. ERSI [http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?topicname=geometrical\\_interval](http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?topicname=geometrical_interval). Accessed 15 July 2017.
- Arribas-Bel, D. (2014). Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography*, 49, 45–53.
- Brown, B., Yamada, I., Smith, K., Zick, C., Kowaleski-Jones, L., & Fan, J. (2009). Mixed land use and walkability: Variations in land use measures and relationships with BMI, overweight, and obesity. *Health and Place*, 15(4), 1130–1141.
- Butler, S. M., Black, D. R., Blue, C. L., & Gretebeck, R. J. (2004). Change in diet, physical activity, and bodyweight in female college freshman. *American Journal of Health Behavior*, 28(1), 24–32.
- Centers for Disease Control and Prevention. (2016). [http://www.paho.org/hq/index.php?option=com\\_contentandview=articleandid=9866andItemid=41164andlang=en](http://www.paho.org/hq/index.php?option=com_contentandview=articleandid=9866andItemid=41164andlang=en). Accessed 1 January 2017.
- Chang, V., & Christaki, N. (2003). Self-perception of weight appropriateness in the United States. *American Journal of Preventive Medicine*, 24(4), 332–339.
- Chen, X., & Yang, X. (2014). Does food environment influence food choices? A geographical analysis through “tweets”. *Applied Geography*. doi:10.1016/j.apgeog.2014.04.003.
- Coomes, E., Jones, A. P., & Hillsdon, M. (2010). The relationship of physical activity and overweight to objectively measured green space accessibility and use. *Social Science and Medicine*, 70(6), 816–822.
- Desmond, S., Price, J., Gray, N., & O’Connell, K. (1986). The etiology of adolescents’ perceptions of their weight. *Journal of Youth and Adolescence*, 15(6), 461–474.
- Donnelly, J. E., Blair, S. N., Jakicic, J. M., Manore, M. M., Rankin, J. W., & Smith, B. K. (2009). American College of Sports Medicine Position Stand. Appropriate physical activity intervention strategies for weight loss and

- prevention of weight regain for adults. *Medicine and Science in Sports and Exercise*, 41(2), 459–471.
- Drewnowski, A., & Rolls, B. J. (2012). Obesity treatment and prevention: New directions. *73rd Nestle Nutrition Institute Workshop*. Carlsbad, CA.
- Fan, M., & Jin, Y. (2015). The effects of weight perception on adolescents' weight-loss intentions and behaviors: Evidence from the youth risk behavior surveillance survey. *International Journal of Environmental Research and Public Health*, 12(11), 14640–14668.
- Fischer, M. M., Leitner, M., & Stauffer-Steinnocher, P. (2001). Spatial point pattern analysis—Some useful tools for analysing locational data. In H. Wohlschlägl (Ed.), *Geographischer Jahresbericht aus Österreich*, 58, 49–66. Vienna: Vienna University, Institute for Geography and Regional Research.
- Frank, L. D., Sallis, J. F., Conway, T. L., Chapman, J. E., Saelens, B. E., & Bachman, W. (2006). Many pathways from land use to health: Associations between neighborhood walkability and active transportation, body mass index, and air quality. *Journal of the American Planning Association*, 72(1), 75–87.
- Ghimire, R., Green, G. T., Ferreira, S., Poudyal, N. C., & Cordeiro, H. K. (2015). Green space and adult obesity prevalence in the United States. *Southern Agricultural Economics Association's 2015 Annual Meeting*. <http://ageconsearch.umn.edu/bitstream/196812/2/green%20space%20obesity.pdf>. Accessed 3 March 2017.
- Ghosh, D., & Guha, R. (2013). What are we 'tweeting' about obesity? Mapping tweets with topic modeling and Geographic Information System. *Cartography and Geographic Information Science*, 40(2), 90–102.
- Hillsdon, M., Panter, J., Foster, C., & Jones, A. (2006). The relationship between access and quality of urban green space with population physical activity. *Public Health*, 120(12), 1127–1132.
- Hoehner, C. M., Brennan Ramirez, L. K., Elliott, M. B., Handy, S. L., & Brownson, R. C. (2005). Perceived and objective environmental measures and physical activity among urban adults. *American Journal of Preventive Medicine*, 28(S2), 105–116.
- Jansenberger, E. M., & Stauffer-Steinnocher, P. (2004). Dual kernel density estimation as a method for describing spatio-temporal changes in the upper Austrian food retailing market, presented at *7th AGILE Conference on Geographic Information Science*. Heraklion, 2004, Greece. [https://agile-online.org/conference\\_paper/cds/agile\\_2004/papers/6-2-3\\_jansenberger.pdf](https://agile-online.org/conference_paper/cds/agile_2004/papers/6-2-3_jansenberger.pdf). Accessed 16 July 2017.
- Lein, J. K. (2012). *Environmental sensing: Analytical techniques for Earth observation* (p. c2012). New York: Springer.
- Levine, N. (2013). Kernel density interpolation. *Office of Justice Programs*. <https://www.nij.gov/topics/technology/maps/documents/crimestat-files/CrimeStat%20IV%20Chapter%2010.pdf> Accessed 16 July 2017.
- Levine, N. (2015). CrimeStat: A spatial statistics program for the analysis of crime incident locations (v 4.02). *Ned Levine & Associates, Houston, Texas, and the National Institute of Justice*, Washington, D.C.
- Liu, J., Bennett, K. J., Harun, N., & Probst, J. C. (2008). Urban-rural differences in overweight status and physical inactivity among US children aged 10–17 years. *The Journal of Rural Health*, 24(4), 407–415.
- Maas, J., Verheij, R., Spreeuwenberg, P., & Groenewegen, P. (2008). Physical activity as a possible mechanism behind the relationship between green space and health: A multi-level analysis. *BioMed Central Public Health*. doi:10.1186/1471-2458-8-206.
- McGee, D. L. (2005). Body mass index and mortality: A meta-analysis based on person-level data from twenty-six observational studies. *Annals of Epidemiology*, 15(2), 87–97.
- Michimi, A., & Wimberly, M. C. (2012). Natural environments, obesity, and physical activity in nonmetropolitan areas of the United States. *Journal of Rural Health*, 28(4), 398–407.
- Multi-Resolution Land Characteristics Consortium (MRLC). National land cover database 2011. (2016). <http://www.mrlc.gov/nlcd2011.php>. Accessed 1 January 2017.
- National Institute of Health. National Heart, Heart, and Blood Institute. (2012). *How are overweight and obesity diagnosed?* <https://www.nhlbi.nih.gov/health/health-topics/topics/obe/diagnosis>. Accessed 1 January 2017.
- Paul, M., & Dredze, M. (2011). You are what you tweet: Analyzing twitter for public health. *Proceedings of fifth international AAAI Conference on Weblogs and Social Media*. <https://drive.google.com/drive/u/0/folders/0B1GsZF0H-pDcVWQxOV9JLXVUYVk>. Accessed 6 February 2017.
- Robillard, J., Johnson, T., Hennessey, C., Beattie, B., & Illes, J. (2013). Aging 2.0: Health information about dementia on Twitter. *PLoS ONE*. doi:10.1371/journal.pone.0069861.
- Silva, M. N., Vieira, P. N., Coutinho, S. R., Minderico, C. S., Matos, M. G., Sardinha, L. B., et al. (2010). Using self-determination theory to promote physical activity and weight control: A randomized controlled trial in women. *Journal of Behavioral Medicine*, 33(2), 110–122.
- Spiegel, A. M., & Alving, B. M. (2005). Executive summary of the strategic plan for National Institutes of Health Obesity Research. *The American Journal of Clinical Nutrition*, 82(1 Suppl), 211S–214S.
- Thornton, L. E., Pearce, J. R., & Kavanagh, A. M. (2011). Using Geographic Information Systems (GIS) to assess the role of the built environment in influencing obesity: A glossary. *International Journal of Behavioral Nutrition and Physical Activity*. doi:10.1186/1479-5868-8-71.
- Twitter. (2016). *Twitter statistics*. <http://www.statisticbrain.com/twitter-statistics/>. Accessed 6 February 2017.
- US DHHS (U.S. Department of Health and Human Services). (2001). *The Surgeon General's call to action to prevent and decrease overweight and obesity*. Rockville, MD: U.S. Department of Health and Human Services, Public Health Service, Office of the Surgeon General.
- US Forest Service. (2015). *Discover camping spots, national park and forests, tours and more*. <https://www.recreation.gov/>. Accessed 1 January 2017.
- US Forest Service. (n.d.). *De Soto Ranger District*. <https://www.fs.usda.gov/detail/mississippi/about-forest/districts/?cid=STELPRDB5209588>. Accessed 1 January 2017.
- Wang, Z., Ye, X., & Tsou, M. (2016). Spatial, temporal, and content analysis of Twitter for wildfire hazards. *Natural Hazards*, 83(1), 523–540.
- Wharton, C., Adams, T., & Hampl, J. (2008). Weight loss practices and body weight perceptions among US college

- students. *Journal of American College Health*, 56(5), 579–584.
- Williams, G. C., Grow, V. M., Freedman, Z. R., Ryan, R. M., & Deci, E. L. (1996). Motivational predictors of weight loss and weight-loss maintenance. *Journal of Personality and Social Psychology*, 70(1), 115–126.
- Wing, R., & Hill, J. (2001). Successful weight loss maintenance. *Annual Review Nutrition*, 21, 323–341.
- Wolch, J., Jerrett, M., Reynolds, K., McConnell, R., Chang, R., Dahmann, N., et al. (2011). Childhood obesity and proximity to urban parks and recreational resources: A longitudinal cohort study. *Health and Place*, 17(1), 207–214.
- World Health Organization. (2016). *Diabetes*. <http://www.who.int/mediacentre/factsheets/fs312/en/>. Accessed 3 March 2017.