Manuscript

Urban Analytics and City Science

EPB: Urban Analytics and City Science
2022, Vol. 0(0) 1–17

DOI: 10.1177/23998083221100827
journals.sagepub.com/home/epb
SAGE

Developing urban biking typologies: Quantifying the complex interactions of bicycle ridership, bicycle network and built environment characteristics

Ben Beck
School of Public Health and Preventive Medicine, Monash University, Melbourne, VIC, Australia

Meghan Winters
Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada

Trisalyn Nelson
Department of Geography, UC Santa Barbara, Santa Barbara, CA, USA

Chris Pettit and Simone Z Leao
City Futures Research Centre, University of New South Wales, Kensington, NSW, Australia

Meead Saberi
School of Civil and Environmental Engineering, University of New South Wales, Kensington, NSW, Australia

Jason Thompson, Sachith Seneviratne and Kerry Nice
Melbourne School of Design, The University of Melbourne, Melbourne, VIC, Australia

Mark Stevenson
Melbourne School of Design, The University of Melbourne, Melbourne, VIC, Australia; Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, VIC, Australia

Abstract
Extensive research has been conducted exploring associations between built environment characteristics and biking. However, these approaches have often lacked the ability to understand the interactions of the built environment, population and bicycle ridership. To overcome these limitations, this study aimed to develop novel urban biking typologies using unsupervised machine learning methods. We conducted a retrospective analysis of travel surveys, bicycle infrastructure

Corresponding author:
Ben Beck, School of Public Health and Preventive Medicine, Monash University, 553 St Kilda Road, Melbourne, VIC 3004, Australia.
Email: ben.beck@monash.edu

and population and land use characteristics in the Greater Melbourne region, Australia. To develop the urban biking typology, we used a k-medoids clustering method. Analyses revealed 5 clusters. We highlight areas with high bicycle network density and a high proportion of trips made by bike (Cluster 1; reflecting 12% of the population of Greater Melbourne, but 57% of all bike trips) and areas with high off-road and on-road bicycle network length, but a low proportion of trips made by bike (Cluster 4, reflecting 23% of the population of Greater Melbourne and 13% of all bike trips). Our novel approach to developing an urban biking typology enabled the exploration of the interaction of bicycle ridership, the bicycle network, population and land use characteristics. Such approaches are important in advancing our understanding of bicycling behaviour, but further research is required to understand the generalisability of these findings to other settings.

Keywords
bicycling, cycling, transportation, spatial variation, typologies

Introduction

Spatial variation in bicycle ridership exists both within and between cities (Beck et al., 2021b; Branion-Calles et al., 2021; Buehler and Pucher, 2012; Firth et al., 2021; Goel et al., 2021; Stevenson et al., 2016). Understanding this variation is important for benchmarking and identifying inequities in participation and access to safe and connected bicycling infrastructure for all ages and abilities. It is well-known that the presence and quality of bicycling infrastructure has a significant impact on bicycle ridership (Pucher and Buehler, 2017), and there is considerable scope for increases in bicycle ridership participation when high-quality and connected infrastructure is provided (Pearson et al., 2022). In Melbourne, Australia, only 1.7% of trips are made by bike (Beck et al., 2021b), despite 78% of the population expressing an interest in riding a bike, but only in the presence of infrastructure that separates riders from motor vehicles (Pearson et al., 2022). Advancing knowledge on the role of the built environment in supporting cycling is necessary to increase bicycling mode share.

Numerous studies have explored the association between the bicycle network, population and land use characteristics, and measures of bicycle ridership (such as mode share) (Buehler and Pucher, 2012; Chen et al., 2017; Christiansen et al., 2016; Kamel and Sayed, 2020; Nelson et al., 2021b; Schoner and Levinson, 2014; Yang et al., 2019; Winters et al., 2010). Measures of bicycle network characteristics that have been previously demonstrated to be associated with bicycle ridership include the length of the bicycle network, network centrality, intersection density and topography (Osama et al., 2017; Kamel and Sayed, 2020; Schoner and Levinson, 2014; Yang et al., 2019). Population and land use characteristics that have been previously demonstrated to be associated with bicycle ridership include population or residential density, and land use mix (Christiansen et al., 2016; Nelson et al., 2021b; Yang et al., 2019). These studies have been important in quantifying the importance of built environment characteristics and connected bicycle networks in enabling cycling. However, these approaches have been limited in their ability to quantify the complex interactions of these factors. In the vast majority of these studies, multivariable regression models (such as lognormal regression, generalised linear mixed models and logistic regression) are used to explore the association between bicycle ridership, built environment and bike network characteristics. Using these approaches yields a single coefficient for each independent variable (often with a confidence or credible interval) that measures the association with a dependent variable (a measure of bicycle ridership). Often, these coefficients are interpreted as an effect that is consistent over a large area, despite having wide confidence or credible intervals. This has limited our understanding of interactions between bicycle ridership, built environment and bicycle network characteristics across space. Quantifying this complexity is needed to identify areas where previously established associations do not hold true, such as areas where there is a well-established bicycle network, but low ridership. One potential opportunity to quantify this complexity is through the use of clustering approaches that bring together bicycle ridership, built environment and bicycle network characteristics in a single model.

Quantifying the complexity of cities has been achieved through approaches such as the development of typologies. As an example, the urban form is often described as the physical configuration of a city that includes the complex relationships between elements such as land use patterns, population and housing densities, infrastructure and amenities, and transport and communication networks (Abrantes et al., 2019). Urban typologies have been developed to simplify this complexity and group areas into similar clusters. They have been used to compare city forms (Thompson et al., 2020; Schirmer and Axhausen, 2016), develop live ability indices (Higgs et al., 2019) and evaluate transport networks (Oke et al., 2019), for example.

Clustering approaches, such as those used to develop urban form typologies, have been used in cycling research as applied to classifying mode share (Goel et al., 2021), bike share schemes (Ma et al., 2019) and bicycle crashes (Sivasankaran and Balasubramanian, 2020). Previous research has also developed bicyclist/population typologies based on individual characteristics, such as the Geller Typology, which characterises individuals based on their confidence in riding in a variety of road/street/path configurations (Dill and McNeil, 2013; Hosford et al., 2020; Pearson et al., 2022). However, to our knowledge, typologies have not previously been used to measure the complexity of the interaction between bicycle mode share, bicycle infrastructure and population and land use characteristics. Therefore, this study aims to develop novel urban biking typologies using unsupervised machine learning methods to classify the complex interaction of bicycle ridership, the bicycle network, population and land use characteristics.

Methods

Study design

We conducted a retrospective analysis of travel surveys, bicycle infrastructure and population and land use characteristics in the Greater Melbourne region, Australia.

Setting

The State of Victoria, Australia, has a population of 6.7 million people of which 67% reside in the Greater Melbourne area (Australian Bureau of Statistics, 2020a). The Australian Bureau of Statistics (ABS) define seven hierarchical classifications of functional areas in Australia, from mesh blocks (the smallest unit) to the country level (the largest unit). Within these is a functional area known as Statistical Areas Level 2 (SA2), which are medium-sized general-purpose areas, reflecting a community that interacts together socially and economically. Statistical Areas Level 2 generally have a population range of 3,000–25,000 persons, with an average of approximately 10,000 persons. For this study, analyses were restricted to the 309 SA2 areas within Greater Melbourne, reflecting an area of 9986 km^2.

Data sources

Bicycle ridership data. Travel survey data were captured through the Victorian Integrated Survey of Travel & Activity (VISTA), coordinated by the Victorian Department of Transport. VISTA is the most robust data available on cycling in the study area. We used data from three waves (2012–14, 2014–16 and 2016–18) of the VISTA. VISTA is a survey of day-to-day travel conducted in the Greater Melbourne area and in a single regional centre in Victoria. Victorian Integrated Survey of Travel & Activity randomly selects households to complete a travel diary for a single specified day, using a stratified, clustered sampling methodology. The survey and resulting data are then weighted to generate population-representative data. Since 2012, 16,000 households and 66,0000 people have contributed to the VISTA survey. Further information on the VISTA is provided elsewhere (Beck et al., 2021b). In this study, we employed a set of combined weights that use the full data set from 2012 to 2018 to produce statistics weighted to the 2017–18 population. Weights were applied to the SA2 in which the trip originated, and therefore, data reflect where trips commenced and not trip routes or destinations. Unless otherwise specified, data reflect trips made within Greater Melbourne on an average day across the study period. Eligibility for inclusion in this study were participants aged 18 years and older, and trips that had trip origins and destinations within the Greater Melbourne region. The metric of bicycle ridership reported in this study is the proportion of all trips (all modes) made by bike within a SA2 area (based on trip origin).

Bicycle network data. We used OpenStreetMap (OSM) data to characterise bicycle infrastructure in the study region. We captured infrastructure at a single time-point, which was the final year of the study period (2018). 2018 OSM data were downloaded for the Greater Melbourne region from Geofabrik (2021). Bicycle infrastructure is coded by OSM contributors according to the OSM Wiki (OpenStreetMap, 2021) and stratified into: on-road bicycle lanes, protected on-road bicycle lanes, and off-road paths (off-road dedicated bicycle path, off-road shared path (shared with pedestrians), and footways where cycling is legal). Further information on this method is described previously (Beck et al., 2021b). Aggregate measures for bicycle network data were calculated for each SA2.

Population and land use characteristics. Population characteristics (population density) were sourced from the ABS (Australian Bureau of Statistics, 2020b). Land use characteristics were derived from ABS Mesh Blocks. Mesh Blocks are the smallest geographical area defined by the ABS and broadly identify land use (e.g. residential, commercial, primary production and parks) (Australian Bureau of Statistics, 2016). Aggregate measures of land use mix were then derived for each SA2.

Procedures

The development of an urban biking typology was informed by previous research that explored the association between cycling network characteristics and bicycle ridership (Beck et al., 2021a). Specifically, measures of the length of the bicycle network, network centrality and topography have previously been demonstrated to be associated with bicycle ridership (Kamel and Sayed, 2020; Osama et al., 2017; Schoner and Levinson, 2014; Yang et al., 2019). Other population and land use characteristics, such as population density and land use mix, have also previously been demonstrated to be associated with bicycle ridership (Christiansen et al., 2016; Nelson et al., 2021b; Yang et al., 2019).

Similar to prior studies that have developed typologies/indices (Higgs et al., 2019; Winters et al., 2010), the final set of indicators chosen for inclusion in the urban biking typology reflected a balance of parsimony and pragmatism. We acknowledge that the included indicators do not capture the full complexity of bicycle ridership and urban form, but represent key attributes that have previously been associated with bicycle ridership.

For measures of bicycle ridership, we selected the proportion of trips that were made by bike, per SA2. For characteristics of the bicycle network in each SA2, we drew from our previous research to calculate measures of bicycle network length (stratified by bicycle infrastructure that was off-road and on-road), network centrality (degree centrality), and topography (Beck et al., 2021a). To calculate degree centrality, we firstly characterised the OSM bicycle network data as a graph, where the links represent the bicycle network infrastructure (e.g. off-road path or on-road bike lane) and the nodes represent the intersections between network links (e.g. street and bicycle network links). To process the OSM data, we 'contracted' the network to remove network artefacts and ensure that the networks contained only edges that directly connected junctions. Topography was measured using the average weighted slope of the bicycle network within the SA2. Elevation data were sourced from the Victorian Government 'Vicmap Elevation' product, which includes a Digital Elevation Model (DEM) at 10 m grid resolution (Victorian Department of Environment Land Water and Planning, 2021) and applied to the OSM network using the 'slopes' package in R (ITS Leeds, 2020). Further information on these methods is described previously (Beck et al., 2021a).

To quantify land use mix in each SA2, we used the Entropy Index (Song et al., 2013). The Entropy index, which varies from 0 to 1, takes into account the relative percentage of land use types within an area. Therefore, higher levels of Entropy correspond with greater land use mix. Similar to prior research (Christian et al., 2011; Christiansen et al., 2016), we selected a subset of land use types that have been shown to be associated with cycling (Christiansen et al., 2016; Chen et al., 2017; Cui et al., 2014): residential, commercial and education. To calculate the Entropy Index, the distribution of land use types in each SA2 was determined through the use of ABS Mesh Block land use classifications. The Entropy Index was then calculated using the following equation (Song et al., 2013)

$$ENT = \frac{-\left[\sum_{j=1}^{k} P^j \ln(P^j)\right]}{\ln(k)}$$

where ENT is the Entropy Index; P^j is the percentage of each land use type j in the area (with the denominator being the summed area for land use classes of interest); and k is the number of land use types.

Analysis

To develop the urban biking typology, we used a commonly employed unsupervised machine learning clustering approach known as Partitioning Around Medoids (PAM) (Kaufman and Rousseeuw, 2009), also simply referred to as k-medoids (Schubert and Rousseeuw, 2021). A variety of clustering methods exist in the literature, including hierarchical clustering, latent class clustering, k-means clustering and k-medoid clustering. K-means is commonly used in the urban and biking literature (Goel et al., 2021; Ma et al., 2019). However, it has been shown to be sensitive to noise and outliers (Velmurugan and Santhanam, 2010), and as a result, k-medoid is recommended over k-means (Arora and Varshney, 2016). K-medoid applies clustering algorithms to find k clusters in n objects by firstly identifying representative objects (the Medoids) for each cluster (the representative example of the members within that cluster), and then assigns each object to the nearest Medoid (Velmurugan and Santhanam, 2010). Based on a review of the literature related to key bicycle ridership, built environment and bicycle network characteristics, the following indicators were included: the proportion of trips made by bike, population density, land use entropy, network density, off-road bicycle network length, on-road bicycle network length and average weighted slope. Data for each measure were standardised and z-scores are presented. The number of clusters was guided by use of 'NbClust' (an R package for determining the relevant number of clusters in a dataset) (Charrad et al., 2014) and the authors' knowledge of the study area. There are a wide variety of indices that have been proposed in the literature to guide the selection of the optimal number of clusters, such as the Silhouette Index and the Gap Index, and these indices often provide differing numbers of clusters (Charrad et al., 2014). 'NbClust' overcomes these limitations by calculating 30 relevant indices and providing a summary of the recommended number of clusters for each index; it is recommended that the optimal number of clusters is based on a majority rule (the highest frequency number of clusters as determined by the 30 indices) (Charrad et al., 2014). A sensitivity analysis was conducted to evaluate the impact on the urban bicycle typology of applying weights to the SA2 in which the trip was completed (the destination), as opposed to the SA2 in which the trip originated. To summarise the characteristics of each cluster, we report the proportion of land area and population captured by each cluster, and report the following measures of bicycle ridership drawn from the VISTA travel survey data: the proportion of all bike trips that are captured within each cluster, the proportion of bike trips made by females, and the proportion of bike trips made for work and recreational purposes.

Data were prepared using the statistical software package R v4.0.3 (R Core Team, 2021) and the integrated development environment RStudio (RStudio 2020, Boston, MA, USA), using the 'survey', 'srvyr', 'dodgr', 'igraph', 'slopes', 'cluster', 'factoextra', 'NbClust' and 'tmap' libraries.

Results

On an average day in Greater Melbourne, there were 180,393 trips made by bicycle, reflecting 1.7% of all trips. The median proportion of trips made by bicycle per SA2 was 0.74% (Q1: 0.26%; Q3: 1.61%) with the highest proportion of trips concentrated in the inner Melbourne region (Figure 1).

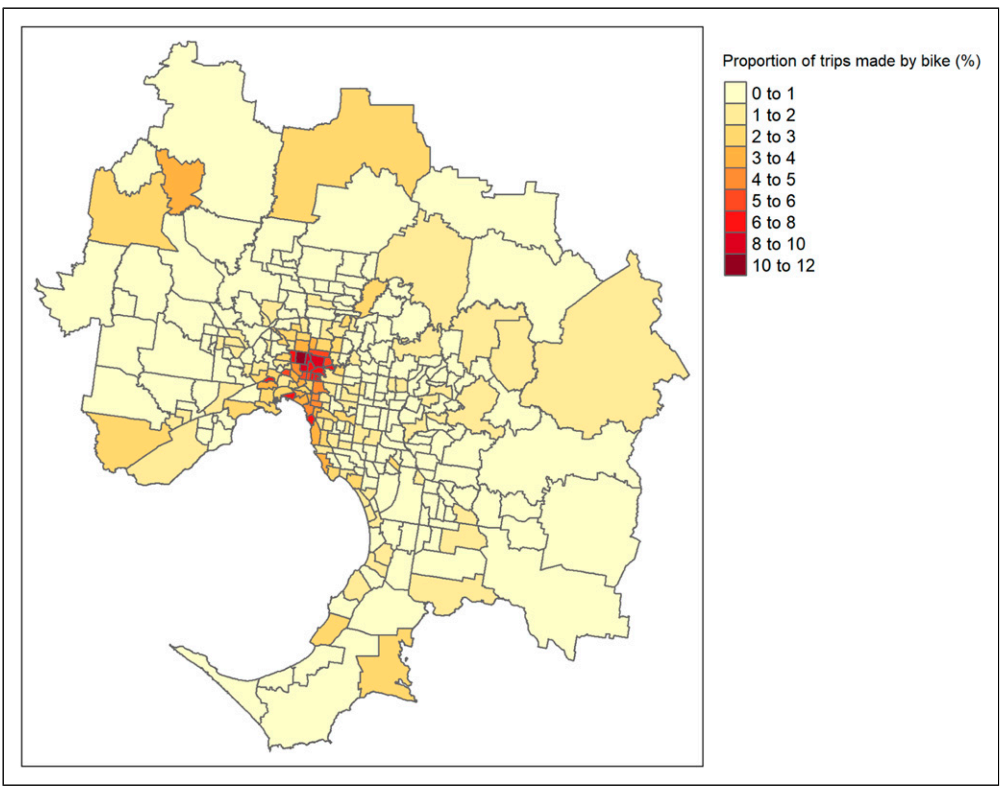

Figure 1. Proportion of all trips that were made by bike (per SA2).

The median length of off-road bicycle infrastructure per SA2 was 5.98 km (Q1: 2.51 km; Q3: 10.47 km) and the median length of on-road bicycle infrastructure per SA2 was 2.33 km (Q1: 0.23 km; Q3: 5.78 km). Network characteristic data are provided in Table 1 and Figure 2. Median land use entropy was 0.24 (Q1: 0.15; Q3: 037) (Table 1 and Figure 3).

K-medoid analysis identified 5 clusters (Figures 4 and Figure 5):

- Cluster 1: High population density, high bicycle network density and a high proportion of trips made by bicycle
- Cluster 2: Average population density, average on-road bicycle network length and below average proportion of trips made by bicycle
- Cluster 3: Below average population density, low land use entropy, below average bicycle network density and low proportion of trips made by bicycle
- Cluster 4: Below average population density, high off-road and on-road bicycle network length, average network density and below average proportion of trips made by bicycle
- Cluster 5: Low population density, low land use entropy, low network density, high average weighted slope and below average proportion of trips made by bicycle.

A map of clusters by SA2 is presented in Figure 5. Cluster 1 was generally observed in inner city areas, Cluster 2 was generally observed in inner urban areas surrounding Cluster 1, Clusters 3 and 4 were generally observed in middle and outer fringe areas, and Cluster 5 was observed exclusively in outer fringe areas. Cluster 1 captured 12% of the population of Greater Melbourne, but 57% of all bike trips (Table 2). Cluster 2 captured 43% of the population of Greater Melbourne and 26% of all bike trips. The proportion of bike trips made by females varied from 23% in Cluster 3–41% in Cluster 1. The proportion of bike trips made for work purposes varied from 16% in Cluster 5–40% in Cluster 1, while the proportion of bike trips made for recreational purposes varied from 16% in Cluster 1–53% in Cluster 5 (Table 2).

Summary characteristics across SA2s in each cluster are presented in Table 3. Median values of the proportion of trips made by bike varied from 0.05% in Cluster 3 (Q1: 0%, Q3: 0.57%) to 5.39% in Cluster 1 (Q1: 3.61%, Q3: 6.53%). Median network density ranged from 0.05 in Cluster 5 (Q1: 0.02, Q3: 0.17) to 4.1 in Cluster 1 (Q1: 2.61, Q3: 5.15).

A sensitivity analysis was conducted using data on where the trip was completed (the destination), as opposed to where the trip originated. The clustering and spatial distribution of clusters

Table 1. Summary statistics of biking, network, population and land use data across SA2s in Greater Melbourne.

Measure	Greater Melbourne [median (Q1, Q3)]
Bicycle ridership data	
Proportion of trips made by bike (%)	0.74 (0.26, 1.61)
Network data	
Off-road bicycle network length (km)	5.98 (2.51, 10.47)
On-road bicycle network length (km)	2.33 (0.23, 5.78)
Network density	1.19 (0.50, 1.97)
Average weighted slope (%)	1.84 (1.15, 2.84)
Population and land use data	
Population density (persons per square km)	2075 (947, 3010)
Land use entropy	0.24 (0.15, 0.37)

Note: Q1: quartile 1; Q3: quartile 3.

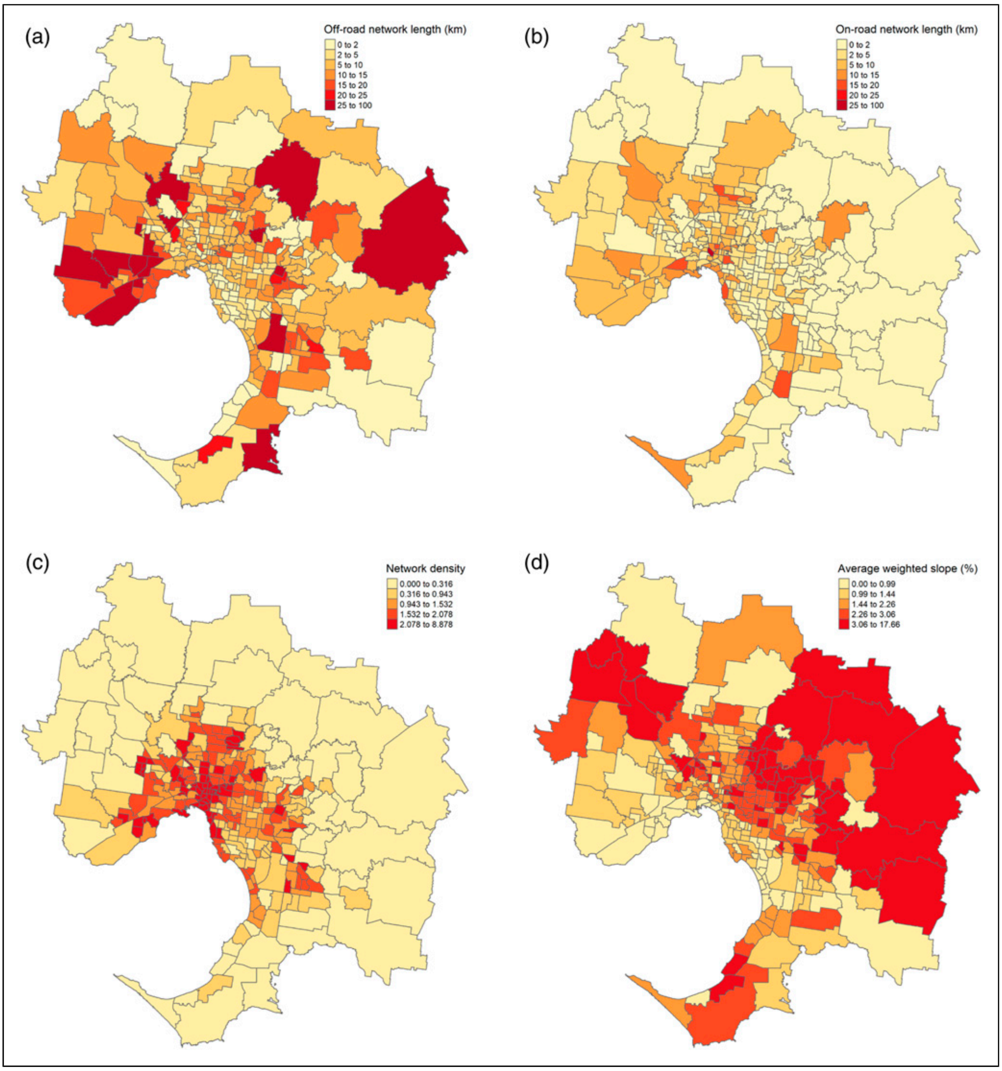

Figure 2. Network characteristics (per SA2). (a) Off-road bicycle network length; (b) On-road bicycle network length; (c) Network density; (d) Average weighted slope.

was similar between the two methods (see Supplementary Material) with 280 of 309 SA2s (91% of all SA2s) having the same cluster assignments between the two methods.

Discussion

In this study, we developed a novel urban biking typology that combined measures of bicycle ridership, bicycle network, and population and land use characteristics, and mapped these clusters across a large metropolitan region. The novel methodological approach that we have employed in this study provides a mechanism to understand the complex interactions between a diversity of factors related to bicycle ridership.

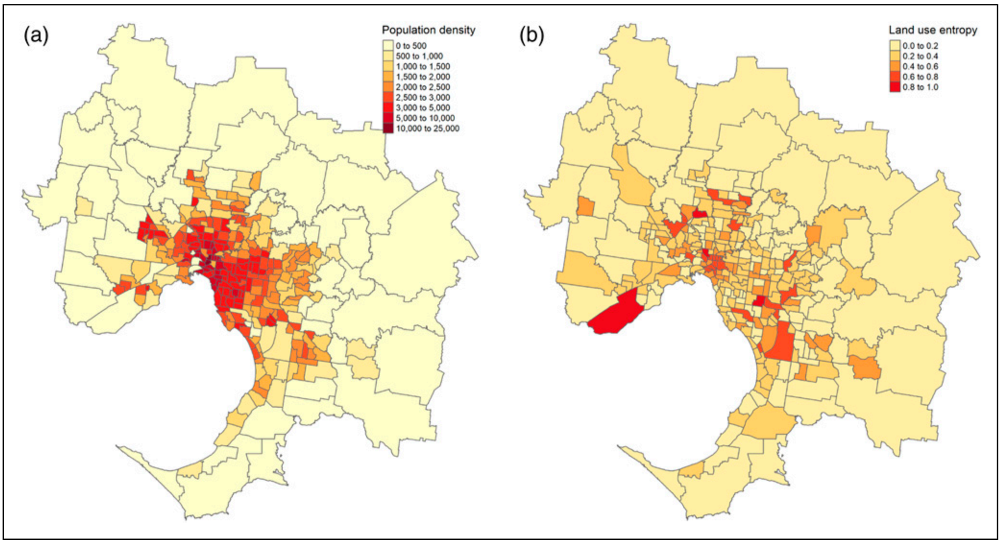

Figure 3. Population and land use characteristics (per SA2). (a) Population density (persons per square km); (b) Land use entropy.

Extensive research has been conducted exploring associations of built environment characteristics and bicycle ridership (Wang et al., 2016; Yang et al., 2019). These studies have demonstrated the importance that bicycle network accessibility, connectivity and comfort, land use mix, population characteristics (including socioeconomic status) and topography have on bicycle ridership (Chen et al., 2017; Christiansen et al., 2016; Nelson et al., 2021b; Winters et al., 2010; Yang et al., 2019). The vast majority of these prior studies have used various forms of regression modelling to explore associations between these built environment and population characteristics, and bicycle ridership, typically generating single coefficients to quantify the association. These approaches are limited in their ability to understand the interactions of built environment, population and cycling characteristics. For example, numerous studies have demonstrated positive associations between bicycle network characteristics (such as length and density) and bicycle ridership (Kamel and Sayed, 2020; Kamel et al., 2020; Osama et al., 2017). However, there may be areas in which there is a relatively high bicycle network length or density, but low bicycle ridership, and conversely areas with low bicycle network length or density and high bicycle ridership (such as Cluster 5). In this study, we overcome the limitations of previous approaches by applying a novel clustering approach that integrates built environment, population and bicycling characteristics.

Our results demonstrate a number of interesting findings. Firstly, Cluster 4 reflects areas with average bicycle network density (and above average on-road and off-road bicycle network length) but a low proportion of trips by bicycle. Understanding the drivers of low bicycle riding in these areas, particularly as many of these areas are in middle and outer urban areas, is needed to advance our understanding of how to enable bicycle riding and how to enhance equity. It may be in these areas that on-road painted bicycle lanes (99.3% of on-road bicycle infrastructure in Melbourne (Beck et al., 2021b)) are inadequate in providing safe and comfortable spaces for bicycle riding; the limitations of on-road painted bicycle lanes are well established in the literature (Beck et al., 2019). It may also be that the distance required to reach destinations of interest by bike exceeds a tipping point, and consideration for support of e-bikes or integration with public transport may assist in increasing bicycle mode share in these areas. Secondly, Cluster 5 reflects areas with low network

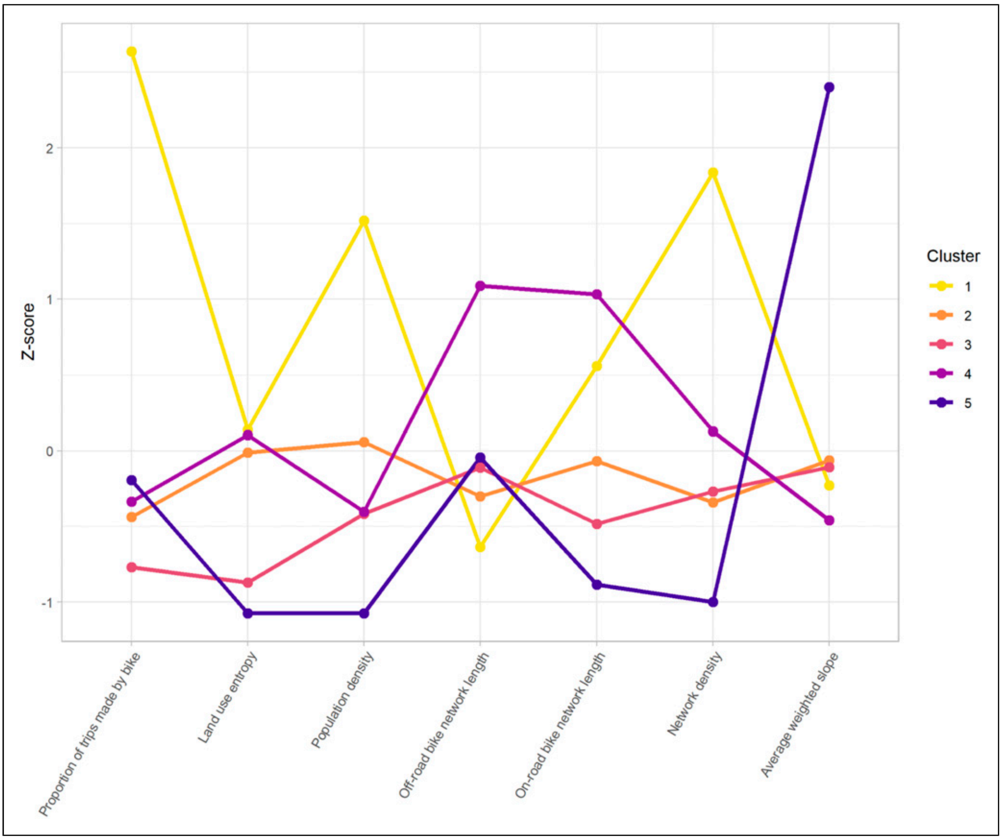

Figure 4. Plot of cluster characteristics with the *y*-axis reflecting the Z-score.

density (but average off-road bicycle network length), high average weighted slope and low population density, but a near-average proportion of trips made by bicycle. This is somewhat surprising as these network, built environment and population characteristics are normally negatively associated with bicycle ridership. It may be that in these areas, a limited but well-connected network of off-road bicycle paths facilitates recreational cycling. This is supported by our finding that 53% of trips in this cluster were for recreational purposes. It is, however, important to note that even an 'average' proportion of trips made by bicycle in Greater Melbourne is still a very low mode share (0.74% of all trips) and significant investment is required to increase bicycle mode share, particularly to ensure equitable access to safe and connected infrastructure across entire metropolitan regions.

The approach developed in this study has multiple potential applications. Firstly, we propose and employ a novel methodological approach for measuring the complex interaction of bicycle ridership, built environment and bike network characteristics. Such an approach has not previously been employed in the literature and we argue that these methods should be considered in future studies that explore bicycle ridership and bicycle networks within and between cities. Secondly, it provides a methodological approach for how bicycle ridership and bicycle infrastructure can be monitored and evaluated within and between cities. As previously described, the association between the quality of bicycle networks and bicycle ridership is not necessarily linear. The typology

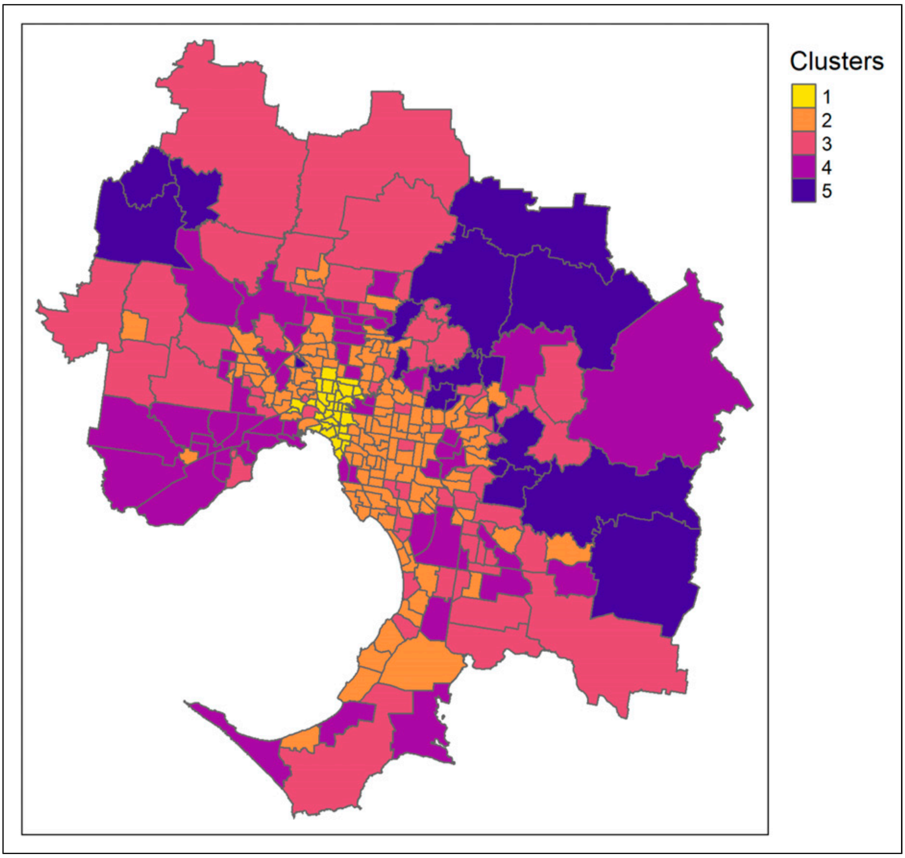

Figure 5. Map of the clusters by SA2.

used in this study enables an understanding of the complex interaction between these factors and subsequently enables the identification of areas where, for example, bicycle ridership is low despite relatively good bicycle network characteristics. This information can then be used to conduct further exploration as to why bicycle ridership is not as high as expected. For the purposes of within or between-city comparisons, these measures provide an opportunity to benchmark performance and explore spatial variation in urban biking typologies. Thirdly, there are opportunities to leverage this typology when considering how to understand behaviour and develop interventions to increase cycling. For example, identifying a discrete set of homogenous areas within cities enables more efficient tailoring of interventions (whether behavioural or infrastructural). It may also assist with sampling when considering capturing the heterogeneity of populations across large geographical areas. It is well recognised that challenges exist in ensuring that samples are either fully or proportionally representative of the populations or activities being considering in spatial and transport planning (Krenn et al., 2011). By reducing the often large number of spatial areas that comprise metropolitan regions, for example, into a discrete set of homogenous clusters, we reduce these sampling biases and provide a sampling framework that can be used to maximise representativeness when collecting mobility data. In our specific application, the urban biking typology enables targeted sampling of mobility data, such as GPS data, to maximise representativeness across a heterogenous area of bicycle ridership, infrastructure, population and land use characteristics.

Table 2. Summary characteristics for each cluster. Note that the bicycle ridership data reflects the distribution of each measure of ridership across the five clusters (i.e. the rows sum to 100%).

Characteristics	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Coverage					
Proportion of land area	1%	11%	40%	24%	24%
Proportion of population	12%	43%	18%	23%	4%
Bicycle ridership data					
Proportion of all bike trips	57%	26%	3%	13%	2%
Proportion of bike trips made by females	41%	24%	23%	33%	25%
Proportion of bike trips made for work purposes[a]	40%	30%	18%	16%	18%
Proportion of bike trips made for recreational purposes[a]	16%	34%	61%	40%	53%

[a]Education, social and other trip purposes not shown.

Given the massive increase in big spatial data relevant to urban environments (Huang and Wang, 2020; Nelson et al., 2021a), the need for tools to aggregate data into typologies is critical for data integration and allows characterization of the city using multiple data sets. Unsupervised machine learning approaches to clustering data is an important approach to characterizing and mapping cities. However, the typologies generated will be dependent on the input variables. Here, we define topologies based on variables known to be important to cycling and that support our goals of stratifying urban settings to optimise cycling related data sampling. A similar approach can be used for other transportation and urban planning applications, and by modifying input variables, the typology can reflect the needs of the application of interest. As such, our approach is a useful framework. However, clustering approaches can be unstable and are sensitive to model inputs (Grubesic et al., 2014). Therefore, generalising this approach to other cities, both in Australia and internationally, requires further development and refinement.

We acknowledge a number of study limitations. For example, survey weights were applied to the SA2 in which the trip originated, and the data presented do not reflect trips that occurred across multiple SA2s. However, our sensitivity analysis showed that the clustering and spatial patterns of clusters were similar when using trip origins and trip destinations. It is also important to acknowledge the variation in size of the spatial areas (SA2s) used in this study; we acknowledge the modifiable areal unit problem and the potential statistical biases this produces. Additionally, due to the relatively low number of trips made by bicycle, we were unable to stratify by trip purpose at the SA 2 level to include it in the clustering (e.g. the proportion of trips that were for recreational purposes). To characterise bicycle infrastructure, we were reliant on OSM data. OpenStreetMap data have been reported to have varied accuracy in international settings (Ferster et al., 2020), and the accuracy of OSM data is unknown in our region. Further, the bicycle network measures that we employed in this study were simplistic due to the absence of bicycle infrastructure data that would enable a more robust evaluation of the accessibility, connectivity and comfort of infrastructure. Further research is required to advance the measurement of these bicycle network parameters. Land use measures were based on ABS Mesh Blocks, which reflect dominant land use and therefore, may not pick up mixed land use at a finer scale. Furthermore, while we used a data-driven approach to the development of typologies, clustering methods are sensitive to model inputs (Grubesic et al., 2014). For the selection of variables, we took an approach that was driven by availability of data and with the approach of achieving a balance of parsimony and pragmatism. We did not consider factors such as access to public transport, social infrastructure, employment and greenspace. Similarly, the

Table 3. Summary characteristics for each cluster reflecting the median, Q1, Q3 and minimum and maximum values across SA2s in each cluster.

Characteristics	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Number					
Number of SA2s	33	132	70	55	19
Bicycle ridership data					
Proportion of trips made by bike (%)	Median: 5.39, Q1: 3.61, Q3: 6.53; [Min: 2.12, Max: 10.84]	Median: 0.97, Q1: 0.42, Q3: 1.47; [Min: 0, Max: 3.22]	Median: 0.05, Q1: 0, Q3: 0.57; [Min: 0, Max: 2.53]	Median: 0.74, Q1: 0.46, Q3: 1.27; [Min: 0, Max: 3.49]	Median: 0.54, Q1: 0, Q3: 1.32; [Min: 0, Max: 3.25]
Network data					
Off-road bicycle network length (km)	Median: 4.21, Q1: 1.88, Q3: 7.16; [Min: 0.11, Max: 15.11]	Median: 5.04, Q1: 2.29, Q3: 7.96; [Min: 0, Max: 15.27]	Median: 6.12, Q1: 2.01, Q3: 9.89; [Min: 0, Max: 18.59]	Median: 17.14, Q1: 11.06, Q3: 25.31; [Min: 1.56, Max: 58.86]	Median: 3.55, Q1: 0.93, Q3: 10.43; [Min: 0, Max: 25.12]
On-road bicycle network length (km)	Median: 6.3, Q1: 4.18, Q3: 11.04; [Min: 0, Max: 25.43]	Median: 2.23, Q1: 0.58, Q3: 3.76; [Min: 0, Max: 11.14]	Median: 0.7, Q1: 0, Q3: 2.19; [Min: 0, Max: 7.14]	Median: 8.08, Q1: 4.58, Q3: 10.64; [Min: 0, Max: 19.18]	Median: 0, Q1: 0, Q3: 0.17; [Min: 0, Max: 4.23]
Network density	Median: 4.1, Q1: 2.61, Q3: 5.15; [Min: 0.78, Max: 8.88]	Median: 1.15, Q1: 0.77, Q3: 1.65; [Min: 0, Max: 3.15]	Median: 0.66, Q1: 0.11, Q3: 1.45; [Min: 0, Max: 3.58]	Median: 1.76, Q1: 0.66, Q3: 2.44; [Min: 0.08, Max: 6.58]	Median: 0.05, Q1: 0.02, Q3: 0.13; [Min: 0, Max: 1.05]
Average weighted slope (%)	Median: 1.56, Q1: 1.19, Q3: 2.12; [Min: 0.3, Max: 3.72]	Median: 2.11, Q1: 1.21, Q3: 2.8; [Min: 0, Max: 5.42]	Median: 1.53, Q1: 0.86, Q3: 2.62; [Min: 0, Max: 4.55]	Median: 1.27, Q1: 0.96, Q3: 2.18; [Min: 0.41, Max: 5.35]	Median: 7.24, Q1: 5.48, Q3: 8.01; [Min: 4.16, Max: 17.66]
Population and land use data					
Population density (persons per square km)	Median: 5658, Q1: 4132, Q3: 7581; [Min: 1896, Max: 20725]	Median: 2544, Q1: 1948, Q3: 3149; [Min: 38, Max: 7078]	Median: 1250, Q1: 130, Q3: 2004; [Min: 0, Max: 3424]	Median: 1411, Q1: 599, Q3: 2111; [Min: 23, Max: 3505]	Median: 124, Q1: 44, Q3: 653; [Min: 0, Max: 2135]
Land use entropy	Median: 0.46, Q1: 0.28, Q3: 0.65; [Min: 0.08, Max: 0.86]	Median: 0.3, Q1: 0.22, Q3: 0.39; [Min: 0.08, Max: 0.9]	Median: 0.11, Q1: 0.06, Q3: 0.16; [Min: 0, Max: 0.24]	Median: 0.28, Q1: 0.19, Q3: 0.38; [Min: 0.05, Max: 0.97]	Median: 0.1, Q1: 0.05, Q3: 0.12; [Min: 0, Max: 0.42]

selection of the number of clusters is a balance between achieving sufficient separation of clusters and complexity. Additionally, we chose not to use a spatially-weighted clustering approach as we were interested in cluster variation in near neighbours. Regardless, the final typology demonstrated substantial spatial clustering. Finally, and as noted above, validation of this approach and the typology developed in other Australian and international cities is needed to confirm generalisability.

Conclusion

In this paper, we present a novel approach to developing an urban biking typology, which classifies the interaction of bicycle ridership, bicycle network, population and land use characteristics. This methodological approach may be used to explore bicycle ridership and bicycle networks within and between cities, to enhance sampling of mobility data, and to inform the development of transport policies and targeted interventions to increase bicycle ridership.

Acknowledgements

The authors would like to acknowledge the assistance provided by the Victorian Department of Transport.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: BB was supported by an Australian Research Council Future Fellowship (FT210100183). The CYCLED (CitY-wide biCycLing Exposure modelling) Study is funded by an Australian Research Council Discovery Project (DP210102089). JT, KN and SS were supported by a National Health and Medical Research Council (NHMRC) and United Kingdom Research and Innovation (UKRI) grant (#1194959).

Ethical approval

Ethical approval for this study was provided by the Monash University Human Research Ethics Committee (Project ID: 29210).

ORCID iDs

Ben Beck https://orcid.org/0000-0003-3262-5956
Chris Pettit https://orcid.org/0000-0002-1328-9830
Sachith Seneviratne https://orcid.org/0000-0001-9094-2736
Kerry Nice https://orcid.org/0000-0001-6102-1292

Supplemental Material

Supplemental material for this article is available online.

References

Abrantes P, Rocha J, Marques da Costa E, et al. (2019) Modelling urban form: a multidimensional typology of urban occupation for spatial analysis. *Environment and Planning B: Urban Analytics and City Science* 46(1): 47–65.

Arora P and Varshney S (2016) Analysis of k-means and k-medoids algorithm for big data. *Procedia Computer Science* 78: 507–512.

Australian Bureau of Statistics (2016) Australian Statistical Geography Standard (ASGS): volume 1 - main structure and greater capital city statistical areas. Available at: https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/1270.0.55.001~July%202016~Main%20Features~Mesh%20Blocks%20(MB)~10012 (accessed 7 February 2021).

Australian Bureau of Statistics (2020a) Australian Bureau of Statistics: data by region. Available at: https://itt.abs.gov.au/itt/r.jsp?databyregion (accessed 30 October 2020).

Australian Bureau of Statistics (2020b) Regional population. Available at: https://www.abs.gov.au/statistics/people/population/regional-population/2018-19 (accessed 22 January 2021).

Beck B, Chong D, Olivier J, et al. (2019) How much space do drivers provide when passing cyclists? Understanding the impact of motor vehicle and infrastructure characteristics on passing distance. *Accident Analysis and Prevention* 128: 253–260.

Beck B, Pettit C, Winters M, et al. (2021a) Association Between Network Characteristics and Bicycle Ridership Across a Large Metropolitan Region. *OSF Preprints*. doi: 10.31219/osf.io/39ke6.

Beck B, Winters M, Thompson J, et al. (2021b) Spatial Variation in Bicycling: A Retrospective Review of Travel Survey Data from Greater Melbourne, Australia. *SocArXiv*. doi: 10.31235/osf.io/78qgf.

Branion-Calles M, Teschke K, Koehoorn M, et al. (2021) Estimating walking and bicycling in Canada and their road collision fatality risks: the need for a national household travel survey. *Prague Medical Report* 22: 101366.

Buehler R and Pucher J (2012) Cycling to work in 90 large American cities: new evidence on the role of bike paths and lanes. *Transportation* 39(2): 409–432.

Charrad M, Ghazzali N, Boiteau V, et al. (2014) NbClust: an R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software* 61(1): 1–36.

Chen P, Zhou J and Sun F (2017) Built environment determinants of bicycle volume: a longitudinal analysis. *Journal of Transport and Land Use* 10(1): 655–674.

Christian HE, Bull FC, Middleton NJ, et al. (2011) How important is the land use mix measure in understanding walking behaviour? Results from the RESIDE study. *International Journal of Behavioral Nutrition and Physical Activity* 8(1): 1–12.

Christiansen LB, Cerin E, Badland H, et al. (2016) International comparisons of the associations between objective measures of the built environment and transport-related walking and cycling: IPEN adult study. *Journal of Transport and Health* 3(4): 467–478.

Cui Y, Mishra S and Welch TF (2014) Land use effects on bicycle ridership: a framework for state planning agencies. *Journal of Transport Geography* 41: 220–228.

Dill J and McNeil N (2013) Four types of cyclists? Examination of typology for better understanding of bicycling behavior and potential. *Transportation Research Record* 2387(1): 129–138.

Ferster C, Fischer J, Manaugh K, et al. (2020) Using OpenStreetMap to inventory bicycle infrastructure: a comparison with open data from cities. *International Journal of Sustainable Transportation* 14(1): 64–73.

Firth C, Branion-Calles M, Winters M, et al. (2021) *Who Bikes? An Assessment of Leisure and Commuting Bicycling from the Canadian Community Health Survey*. Transport Findings.

Geofrabrik (2021) GEOFABRIK//downloads. Available at: https://www.geofabrik.de/data/download.html (accessed 08 January 2021).

Goel R, Goodman A, Aldred R, et al. (2021) Cycling behaviour in 17 countries across 6 continents: levels of cycling, who cycles, for what purpose, and how far? *Transport Reviews* 42: 1–24.

Grubesic TH, Wei R and Murray AT (2014) Spatial clustering overview and comparison: accuracy, sensitivity, and computational expense. *Annals of the Association of American Geographers* 104(6): 1134–1156.

Higgs C, Badland H, Simons K, et al. (2019) The Urban Liveability Index: developing a policy-relevant urban liveability composite measure and evaluating associations with transport mode choice. *International Journal of Health Geographics* 18(1): 1–25.

Hosford K, Laberee K, Fuller D, et al. (2020) Are they really interested but concerned? A mixed methods exploration of the Geller bicyclist typology. *Transportation Research Part F: Traffic Psychology and Behaviour* 75: 26–36.

Huang B and Wang J (2020) Big spatial data for urban and environmental sustainability. *Geo-spatial Information Science* 23(2): 125–140.

ITS Leeds (2020) Slopes package. Available at: https://github.com/ropensci/slopes (accessed 01 June 2021).

Kamel MB and Sayed T (2020) The impact of bike network indicators on bike kilometers traveled and bike safety: A network theory approach. *Environment and Planning B: Urban Analytics and City Science* 2020: 239980832096446.

Kamel MB, Sayed T and Bigazzi A (2020) A composite zonal index for biking attractiveness and safety. *Accident Analysis and Prevention* 137: 105439.

Kaufman L and Rousseeuw PJ (2009) *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: John Wiley & Sons.

Krenn PJ, Titze S, Oja P, et al. (2011) Use of global positioning systems to study physical activity and the environment: a systematic review. *American Journal of Preventive Medicine* 41(5): 508–515.

Ma X, Cao R and Jin Y (2019) Spatiotemporal clustering analysis of bicycle sharing system with data mining approach. *Information* 10(5): 163.

Nelson T, Ferster C, Laberee K, et al. (2021a) Crowdsourced data for bicycling research and practice. *Transport Reviews* 41(1): 97–114.

Nelson T, Roy A, Ferster C, et al. (2021b) Generalized model for mapping bicycle ridership with crowdsourced data. *Transportation Research Part C: Emerging Technologies* 125: 102981.

Oke JB, Aboutaleb YM, Akkinepally A, et al. (2019) A novel global urban typology framework for sustainable mobility futures. *Environmental Research Letters* 14(9): 095006.

OpenStreetMap (2021) OpenStreetMap Wiki. Available at: https://wiki.openstreetmap.org/wiki/Main_Page (accessed 08 January 2021).

Osama A, Sayed T and Bigazzi AY (2017) Models for estimating zone-level bike kilometers traveled using bike network, land use, and road facility variables. *Transportation Research Part A: Policy and Practice* 96: 14–28.

Pearson L, Dipnall J, Gabbe B, et al. (2022) The potential for bike riding across entire cities: quantifying spatial variation in interest in bike riding. *Journal of Transport and Health* 24: 101290.

Pucher J and Buehler R (2017) Cycling towards a more sustainable transport future. *Transport Reviews* 37(6): 689–694.

Schirmer PM and Axhausen KW (2016) A multiscale classification of urban morphology. *Journal of Transport and Land Use* 9(1): 101–130.

Schoner JE and Levinson DM (2014) The missing link: bicycle infrastructure networks and ridership in 74 US cities. *Transportation* 41(6): 1187–1204.

Schubert E and Rousseeuw PJ (2021) Fast and eager k-medoids clustering: O (k) runtime improvement of the PAM, CLARA, and CLARANS algorithms. *Information Systems* 101: 101804.

Sivasankaran SK and Balasubramanian V (2020) Exploring the severity of bicycle–vehicle crashes using latent class clustering approach in India. *Journal of Safety Research* 72: 127–138.

Song Y, Merlin L and Rodriguez D (2013) Comparing measures of urban land use mix. *Computers, Environment and Urban Systems* 42: 1–13.

Stevenson M, Thompson J, de Sá TH, et al. (2016) Land use, transport, and population health: estimating the health benefits of compact cities. *Lancet* 388(10062): 2925–2935.

Thompson J, Stevenson M, Wijnands JS, et al. (2020) A global analysis of urban design types and road transport injury: an image processing study. *The Lancet Planetary Health* 4(1): e32–e42.

Velmurugan T and Santhanam T (2010) Computational complexity between K-means and K-medoids clustering algorithms for normal and uniform distributions of data points. *Journal of Computer Science* 6(3): 363–368.

Victorian Department of Environment Land Water and Planning (2021) Vicmap elevation. Available at: https://www.land.vic.gov.au/maps-and-spatial/spatial-data/vicmap-catalogue/vicmap-elevation (accessed 01 May 2021).

Wang Y, Chau CK, Ng W, et al. (2016) A review on the effects of physical built environment attributes on enhancing walking and cycling activity levels within residential neighborhoods. *Cities* 50: 1–15.

Winters M, Brauer M, Setton EM, et al. (2010) Built environment influences on healthy transportation choices: bicycling versus driving. *Journal of Urban Health* 87(6): 969–993.

Yang Y, Wu X, Zhou P, et al. (2019) Towards a cycling-friendly city: an updated review of the associations between built environment and cycling behaviors (2007–2017). *Journal of Transport and Health* 14: 100613.