Transportation
https://doi.org/10.1007/s11116-025-10599-5

Understanding the impact of city-wide cycling corridors on cycling mode share among different demographic clusters in Greater Melbourne, Australia

Afshin Jafari[1] · Steve Pemberton[1] · Dhirendra Singh[2,3] · Tayebeh Saghapour[1] · Alan Both[1] · Lucy Gunn[1] · Billie Giles-Corti[1,4,5]

Accepted: 20 February 2025

Abstract
In car-dominated cities like Melbourne, Australia, limited data on cyclists' travel patterns and socio-demographic differences complicate understanding of the effectiveness of infrastructure investment interventions aimed at promoting cycling. Recent advancements in city-scale transport modelling enable virtual testing of such interventions. However, the application of agent- and activity-based models for large-scale cycling simulations has been constrained by data and complexity. In this study, we developed a city-scale agent-based simulation model for Greater Melbourne to evaluate changes in travel mode share from cycling infrastructure modifications. We clustered bicycle riders into five demographic groups: Maverick Males, Motivated Adults, Conscientious Commuters, Young Sprinters, and Relaxed Cruisers, estimating mode choice parameters for each group. Using aggregated smartphone application data, we developed a cycling trip routing methodology to incorporate road infrastructure impacts. Results indicated that travel time significantly influences mode choice across all clusters. Cycling infrastructure was crucial for four clusters, and travel cost influenced four clusters. The calibrated model assessed the potential impact of fully implementing Greater Melbourne's strategic cycling corridors, a network of key cycling routes. Simulations suggested an initial 30% increase in cycling use, raising the mode share to approximately 2.6%, indicating a modest overall impact. Further analysis showed that even with full implementation, on average about half of the lengths of the routed bikeable trips would still occur on roads without any cycling infrastructure. This underscores the need to improve infrastructure on both major corridors and minor roads, and to complement these improvements with behavioural interventions.

Keywords Bike use · Bicycle infrastructure · Mode choice model · Active transport · MATSim

Extended author information available on the last page of the article

Published online: 03 April 2025

Introduction

Using sustainable and active modes of transport, such as cycling, is crucial to achieving sustainable urban and transport development goals (Liu et al. 2020). Furthermore, regular cycling provides significant health benefits, including reducing the risk of preventable chronic diseases, which costs Australia $15.6 billion annually (Crosland et al. 2019), and improving mental health and well-being (Garrard et al. 2012).

A well-connected cycling infrastructure is essential to encourage more people to cycle (Furth 2021; Furth et al. 2016; Braun et al. 2016; Buehler and Pucher 2012). By examining daily cycling counts from 736 bicycle counters in 106 European cities, along with announced or completed pop-up bike lanes in these cities, Kraus and Koch (2021) found that, on average, provisional bike lanes implemented during the COVID-19 pandemic resulted in an increase in cycling counts ranging from 11% to 48%. In Lisbon, Portugal, the expansion of the cycling network in inner city areas led to a 3.5-fold increase in cycling between 2016 and 2018 (Félix et al. 2020).

The importance of dedicated cycling infrastructure is even more significant in cities dominated by cars, such as those in Australia and North America, where there is a latent demand for cycling if adequate infrastructure is provided (Pearson et al. 2023). According to the 2011-2019 National Cycling Participation Survey, more than half of Australian households own at least one bicycle; however, the share of cycling in daily travel is only 1.4% (Buehler and Pucher 2021). Evidence indicates that only 1.1% of Australian workers commute by bicycle, although nearly 30% of the working population live within a 30-minute cycle distance from their workplace (Both et al. 2022). In regional city neighbourhoods in Australia, about 60% of workers have access to local employment, yet the vast majority travel to work by private motor vehicles, although 30% of work trips are within 5 km of their home (Giles-Corti et al. 2022).

Several studies have shown that the impact of built-environment factors on an individual's decision to use a bicycle differs significantly between the sex and age groups (Debnath et al. 2021; Garrard 2021; Mitra and Nash 2019; Branion-Calles et al. 2019; Aldred et al. 2017; Heesch et al. 2014; Handy et al. 2010). For example, Goel et al. (2022) found that the share of female bicycle riders is significantly lower than that of male bicycle riders in areas with a total bicycle mode share of less than 7%. Shaw et al. (2020) found that in New Zealand, females cycle less often and for shorter distances than males, however, they use a more diverse set of travel modes each day. The development and improvement of cycling infrastructure are critical in promoting cycling in cities. Wałdykowski et al. (2021) argued that a comprehensive network of cycling routes is more beneficial in increasing cycling participation compared to building subsidiary routes.

In the Australian context where cycling levels are low, based on a cross-sectional study of 1,862 bicycle riders in Queensland, Australia, Heesch et al. (2012) found that males were more likely to cycle for transport than females, and also that females were less willing to travel long distances. In Melbourne, Australia, a survey revealed that most participants (57%) owned a bicycle; however, only 28% of males and 12% of females cycled at least once a week (Pearson et al. 2022).

There is a wide range of factors shown to affect cycling behaviour, including built-environment factors, social enviornment facotrs, individual factors, and the inter-dependencies between these (Trapp et al. 2011). This makes it difficult to predict the potential change in cycling mode share and traffic volume as a result of a change in infrastructure (Ziemke et al. 2019). In cities dominated by cars, such as Melbourne, Australia, there is typically limited information on cyclist preferences and differences between socio-demographic cohorts when it comes to cycling, which makes it even more complicated to design effective cycling infrastructure interventions.

The Victorian Government in Australia in its 2018-2028 Cycling Strategy (Department of Transport 2020) has committed to implement a network of dedicated connected bicycle road infrastructure or cycleways in the Greater Melbourne region, known as Strategic Cycling Corridors (SCC). These corridors represent high-priority routes and link key destinations such as activity centres and the central business district. The SCC encompasses roads with existing high-quality cycleways, roads where existing cycleways will be upgraded, and roads lacking any form of cycleway where new cycleways will be introduced. However, estimating the potential impact of fully implementing these corridors requires an understanding of travel demand in Melbourne, as well as the effect of these infrastructure changes on the travel choices of Melburnians.

City-scale transport models, although traditionally designed to model motorised travel, such as private cars and Public transport (PT), have the potential to be used as a decision support tool for cycling infrastructure interventions (Milakis and Athanasopoulos 2014). In particular, agent-based transport models have great potential to be used for large-scale cycling transport simulation, as they can capture heterogeneous interactions and decision-making processes of travelling agents, where agents represent cyclists, pedestrians, drivers, and other road users. However, to date, agent-based modelling studies on cycling behaviour have been less prevalent and few have considered demographic differences in cycling behaviour. An example is the model developed by Aziz et al. (2018) for New York City to support decisions about investments in walking and cycling infrastructure, which included selected demographic, built environment, and social factors. Transport researchers have also used agent-based models to study more specific types of bike use, such as bike sharing systems (Hollauer et al. 2018) or last-mile distribution using cargo bikes (Llorca and Moeckel 2021) and also to estimate the benefits of specific interventions, such as separated cycling infrastructure (Thompson et al. 2017).

One of the main challenges in building agent-based cycling transport models is the lack of sufficient data for model development and calibration. Travel survey data, collected by various departments of transport, are the most commonly used datasets for building transport models, as they typically include detailed trip-level data for large areas, such as cities or even countries. However, travel surveys often lack information on the routes taken by travellers, which is an important factor in cycling modelling. Even when this information is available, it is typically limited to a small subset of the sample. Furthermore, the limited number of bicycle riders recorded in travel surveys from cities with low cycling mode shares, such as Melbourne, presents another challenge, resulting in the inability to include various demographic and spatial attributes in the simulation model parameter estimation.

In this study, we developed a city-scale, agent-based simulation model with heterogeneous decision-making parameters, designed to simulate scenarios of cycling infrastructure changes in Greater Melbourne, Australia. To overcome data limitations in building cycling simulation models, we proposed a model development workflow that uses a combination of data from universal smartphone applications and travel surveys to create a cycling simulation model. To account for differences among various demographic groups, we used a clustering algorithm to categorise bicycle riders in Melbourne according to their trip characteristics. Subsequently, we estimated the parameters of the simulation model for each cluster. We introduced a popularity-based routing module, informed by smartphone application data, to estimate model parameters that reflect the influence of road infrastructure on agents' cycling behaviours. Furthermore, this study investigates the extent to which the proposed SCC, if fully implemented, could improve access to cycleways for trips within a bikeable distance in Greater Melbourne and examines the potential increase in bicycle use in different demographic cohorts as a result of this infrastructure change intervention.

This study employs a multi-step approach to investigate cycling behaviour and infrastructure in Greater Melbourne, integrating clustering analysis, mode choice modelling, and scenario simulations. The methods used are interconnected to provide a comprehensive understanding of cycling patterns and the impact of potential infrastructure changes. The clustering analysis identifies differences in cycling trip characteristics across demographic cohorts, which align with the agent-based modelling approach where travel distance and time are key factors in decision-making. The mode choice model further examines the impact of new cycling infrastructure on mode share. Together, these components enable the development of simulation models to test the impact of changes in cycling infrastructure using an agent-based model, illustrated through the example of Strategic Cycling Corridors. By integrating insights from these analyses, this paper offers a comprehensive view of the current state of cycling in Greater Melbourne, the heterogeneity in cycling behaviour among Melburnians, and potential opportunities for increased uptake through investments in new cycling infrastructure.

The remainder of the paper is structured as follows: Sect. 2 provides an overview of the current state of bicycle ridership and infrastructure in Greater Melbourne. Section 3 provides an analysis of differences in the characteristics of cycling trips between demographic cohorts using a cluster analysis approach. Section 4 describes the mode choice modelling approach, including the calculation of travel routes for alternative modes and the estimation of mode choice model coefficients. These coefficients are then used to build a multi-modal agent-based and activity-based simulation model, as detailed in Sect. 5. The before-and-after scenario analysis of the full implementation of SCC in Melbourne, including simulation results and a comparison of access to cycleways, is also presented in Sect. 5. Finally, Sect. 6 discusses the implications, contributions, and limitations of our cluster-based city-scale simulation model for cycling.

Analysing cycling trips and infrastructure in Greater Melbourne

To better understand current cycling trips and the travel behaviour of cyclists, in this section, we analyse some of the cycling-related infrastructure attributes in Greater Melbourne, as well as the trip characteristics and individual characteristics of those who ride a bicycle in Greater Melbourne. To do this, we conducted three analyses of current cycling travel patterns, that include: (i) cycleway road coverage, (ii) cycling mode share, (iii) trips within a bikeable distance. We used the OpenStreetMap (OSM) extract for the Greater Melbourne region for November 2023 to map the existing road network and cycling infrastructure in Greater Melbourne. The open-source tool developed by Jafari et al. (2022) was used to create the transport network model of Greater Melbourne from OSM. Figure 1 shows the

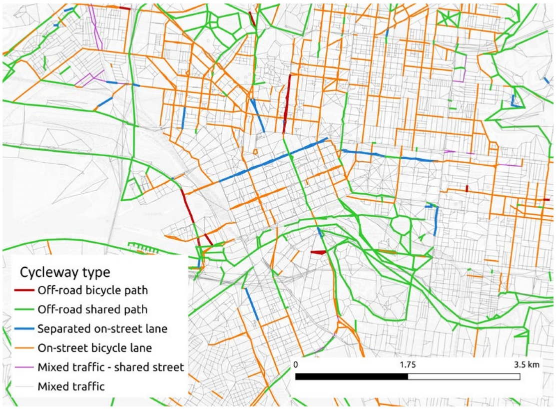


Fig. 1 Road network model with different types of cycleway for Melbourne's inner city area (basemap from OpenStreetMap)

generated transport network and the identified cycleway types for inner Melbourne, which includes Central Business District (CBD) and surrounding areas.

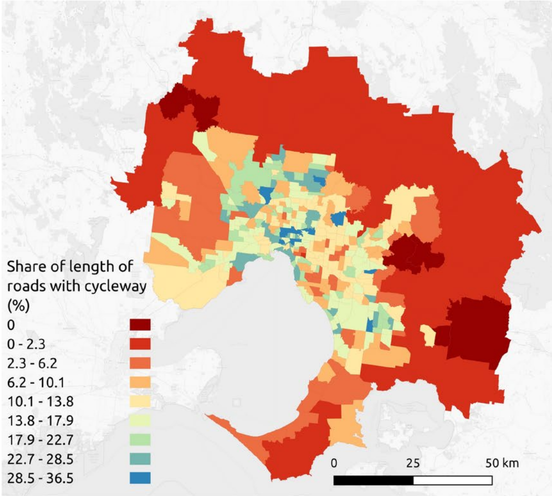
Cycleway road coverage: To understand the coverage of cycleways in Greater Melbourne, we calculated the percentage of road length in each geographic area that includes any type of cycleway. We selected all publicly accessible roads where cycling is permitted, excluding footpaths, highways, and private roads. Each road was divided into smaller segments based on Statistical Area level 2 (SA2) geographical boundaries. SA2s are medium-sized, general-purpose areas defined by the Australian Bureau of Statistics (ABS), with an average population of around 10,000. Coverage percentage was calculated based on the length of roads with any type of cycleway and the total road length in each SA2. As shown in Fig. 2(a), no more than 36% of the length of the roads have any type of cycleway for Melbourne's inner suburbs, which are the areas where cycleways are most concentrated. This percentage falls below 10% in the outer-ring suburbs and growth areas of Greater Melbourne.

Cycling mode share: To examine cycling travel patterns in Greater Melbourne, we used the trips dataset from Victorian Integrated Survey for Travel and Activity (VISTA) for the years 2012 to 2016 within the Greater Melbourne area. For this analysis, we filtered the travel survey trips to include only non-recreational trips[1] that started or ended within the Greater Melbourne region and were undertaken by car, bicycle, public transport (i.e., bus, train, and tram), or walking. Figure 2(b) illustrates the mode share of cycling in the final selected trips dataset, based on their origin SA2 in Greater Melbourne. The figure highlights a significant concentration of cycling in the inner suburbs of Melbourne, where the cycling mode share peaks at approximately 9% in certain areas, while it drops below 1% in more peripheral suburbs.

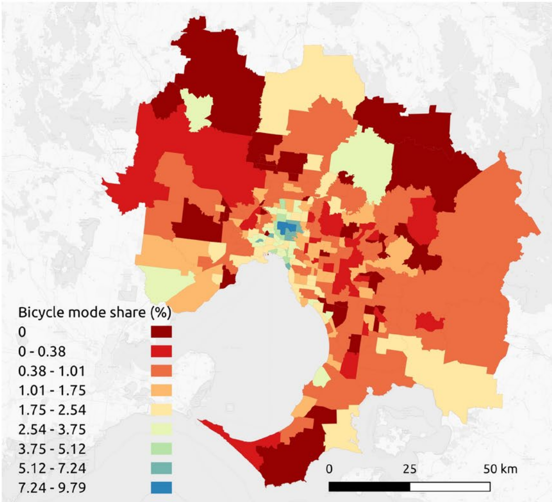
Trips within a bikeable distance: Next, we analysed trips from VISTA that were within a bikeable distance in Greater Melbourne. We chose 5 km as the threshold for a bikeable distance in this analysis, as it is approximately the average distance for cycling trips in Greater Melbourne recorded in VISTA. Additionally, it is the distance that an average bicycle rider can cover at an average speed of 15 km/h in 20 min. Trips from the travel survey were divided into (i) non-work trips, which include trips for education, shopping, health care,

[1] In all the analyses of this study that use VISTA data, trips in the travel survey categorised as recreational, such as walking in a park or running for sport, were excluded.

Fig. 2 Current state of cycling Melbourne in terms of **a** cycleway road coverage percentage and **b** cycling mode share percentage across SA2s in Melbourne (basemap from OpenStreetMap)



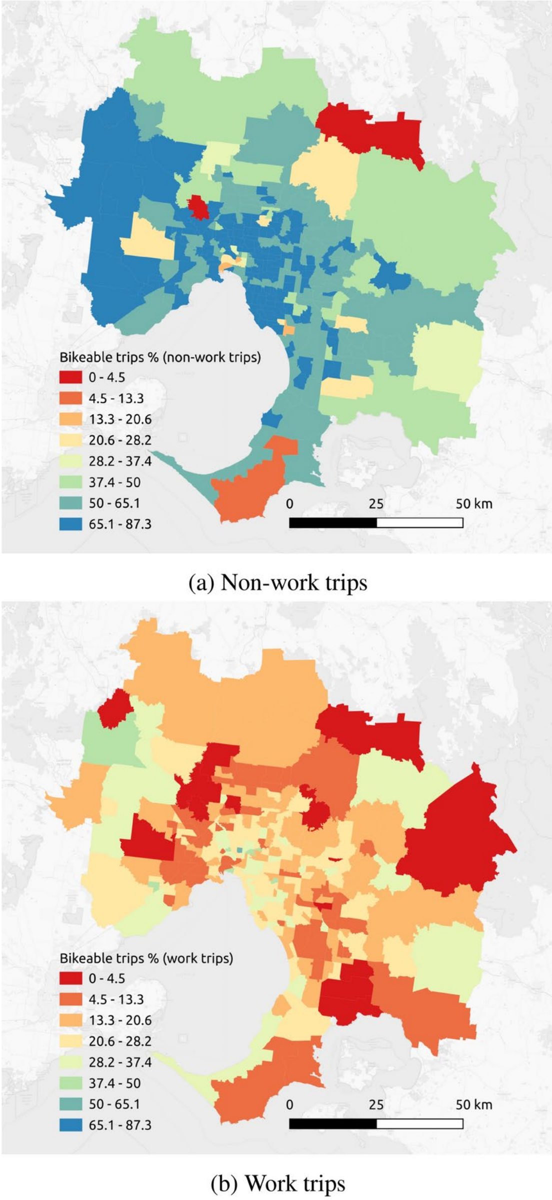
(a) Cycleway road coverage percentage



(b) Cycling mode share percentage

and socialising purposes, and (ii) work trips. For each SA2, percentage of the trips that were less than 5 km was calculated for each trip purpose category. As shown in Fig. 3, there is a considerable difference between work and non-work trips across Greater Melbourne. In most SA2s of Melbourne, about 50% of non-work trips fall within a bikeable distance, with some areas even reaching 80%. This suggests substantial potential for shifting non-work trips to cycling, provided that adequate infrastructure is in place. Figure 3(b), however, shows that a smaller proportion of work trips are within a bikeable distance, and these are mainly concentrated in inner SA2s, where 20 to 30% of work trips are less than 5 km. This is likely due to the higher concentration of jobs within the CBD area of Greater Melbourne.

Fig. 3 Percentages of non-recreational trips from the travel survey that are within a bikeable distance of 5 km for **a** non-work trips and **b** work trips (basemap from OpenStreetMap)



Cluster analysis of bicycle riders in Greater Melbourne

We used cluster analysis to identify groups of travellers, that is, those with at least one trip recorded in the travel survey, who exhibit similar cycling patterns. To do this, we analysed trips from VISTA for the years 2012 to 2016 that were done using bicycle as the main mode of transport and with origin and destination of the trips both within the Greater Melbourne area. The analysis used a two-step cluster approach: initially, bicycle riders from VISTA were categorised into demographic cohorts, followed by cluster analysis of these cohorts based on the characteristics of their cycling trips. This method of clustering demographic cohorts, rather than individuals, was performed to ensure that the final clusters capture differences in the characteristics of cycling trips across different demographic cohorts that are later used to build the simulation model.

Ideally, clustering would consider an extensive list of individual-level factors, such as habits, physical ability, and occupation, in addition to the characteristics of the cycling trip. However, the limited number of cycling trips recorded in VISTA restricted our ability to incorporate all these factors. Incorporating too many factors would lead to creating cohorts with insufficient data. Consequently, we focus only on two demographic attributes of age group and sex (M/F), which are consistently associated with cycling decisions (Goel et al. 2022; Heesch and Turrell 2014; Boulange et al. 2017).

By clustering demographic cohorts rather than individuals, we ensure that the clusters are not only demographically distinct but also behaviourally relevant for the agent-based simulation model (discussed later in Sect. 5.3), where age and gender groups are the main demographic attributes assigned to the agents. Travel distance and time (derived from speed) are the primary factors influencing the agents' decision-making processes, reflecting how agents from different demographic groups value these factors. The primary purpose of the clustering analysis was to support the development of the mode choice model, rather than solely identifying differences in cycling trip behaviour.

For each age-sex cohort C, we used the average travel speed $\bar{v}_C$ and distance $\bar{d}_C$ for cycling trips from the survey data. Cycling speed for each trip was calculated based on reported travel time of that trip and the travel distance. To analyse variations in travel speed and distance between cohorts, each cohort C was plotted as a point on a 2D plane, with travel speed and distance as coordinates.

We normalised travel distance and speed to ensure both variables contributed equally to the clustering process, preventing any one variable from dominating the results due to differing scales. This approach helped produce balanced clusters that accurately reflect distinct cycling behaviours.

We calculated the Euclidean distance (δ) between each cohort pair to quantify their dissimilarity. The Euclidean distance, as a measure of the distance between two points in a 2D space, which, in this case, are the coordinates representing normalised average travel speed and distance for each cohort, allows us to capture how different one cohort is from another based on these two metrics. For each two cohorts $C = [S.A]$ and $C' = [S'.A']$, where S/S' and A/A' denote the sex and age of the members of the cohorts respectively, the Euclidean distance was calculated as illustrated in Eq. 1. The Euclidean distance is calculated by taking the square root of the sum of the squared differences between the corresponding coordinates (normalised average speed and distance) of the two cohorts.

Figure 4 displays the Euclidean distance matrix plot, which visually represents the dissimilarities between all cohort pairs. This plot is valuable as it helps identify which cohorts are more similar or distinct from each other. The presence of both high and low distance values between cohorts suggests that the resulting clusters will capture meaningful distinctions in cycling behaviours.

$$\delta(C, C') = \sqrt{(\bar{v}_C - \bar{v}_{C'})^2 + (\bar{d}_C - \bar{d}_{C'})^2}. \tag{1}$$

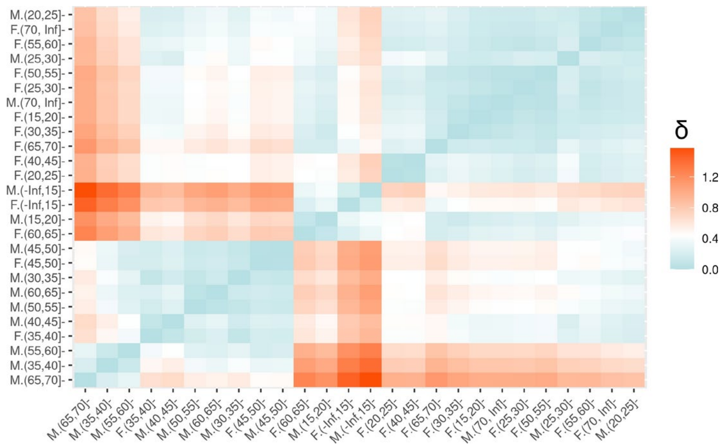


Fig. 4 Matrix of normalised travel speed and the distance difference between demographic cohorts in terms of the Euclidean distance (δ) between the normalised average cycling travel speed and distance for every two cohorts

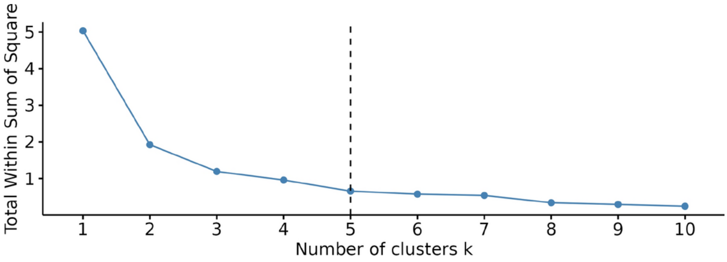


Fig. 5 Total Within-Cluster Sums of Squares values for different number of clusters

The K-means clustering algorithm was used to segment demographic cohorts into a set of K clusters. The number of clusters, K, serves as an input for this algorithm. To determine an appropriate value for K, cluster analysis was performed for $K = 1, ..., 10$, considering cycling travel speed and distance as variables. The most suitable number of clusters was identified by examining the change in total within-cluster variation for different K values, assessed using the within-cluster sum of squares, W_K. To calculate W_K, first, for each data point i within each cluster $k \in K$, the Euclidean distance to the cluster's average, μ_k, was calculated using Eq. 1. Here, μ_k represents a vector of the average travel speed and distance for all data points within the cluster, and the Euclidean distance from data point i to μ_k is represented as $\delta(i, \mu_k)$. W_K was then determined using Eq. 2.

$$W_k = \sum_{k \in K} \sum_{i \in k} (\delta(i, \mu_k))^2. \tag{2}$$

Figure 5 shows W_K for different cluster sizes, $K = 1, ..., 10$, indicating that $K = 5$ is the point after which the increase in K results in a less significant change in the total within-cluster variation value and offers a good balance between the total number of clusters and the differences between them. Thus, $K = 5$ was chosen as the appropriate number of clusters. The final clusters with the sex and age label of each demographic cohort are shown in Fig. 6. These clusters were used as the basis for building the mode choice models of the simulation model.

As indicated in Table 1, the first cluster comprises 20.7% of total travellers and 32.0% of all cyclists in VISTA, consisting entirely of middle-aged and older male bicycle riders. These riders reported longer travel distances and higher cycling speeds compared with other cohorts, which resulted in the label *Maverick Males*. This label was chosen because the average travel distance and speed of those in this cluster push beyond the typical cycling patterns observed in Melbourne. The next cluster includes male and female bicycle riders under 15 years and female cyclists aged 60 to 65 years, representing 21.6% of all travellers and 22.1% of cyclists. This group, which exhibits the shortest travel distances and lowest speeds, was named *Relaxed Cruisers*. The third cluster *Conscientious Commuters*, is the largest in terms of the VISTA population coverage, encompassing 28.0% of the population but only 17.5% of cyclists. It mainly consists of young and middle-aged women and males aged 70 and above. According to Fig. 6, these riders travel shorter distances and at slower speeds than average. The fourth cluster, which includes female cyclists aged 15 to 20 and 30 to 35, as well as male cyclists aged 15 to 30, accounts for 14.9% of the total sample and 14.1% of cyclists, travelling shorter distances but at higher speeds than average. This group is identified as *Young Sprinters*. Lastly, the final cluster, labelled *Motivated Adults*, consisted of 14.9% male and female adults cycled longer distances at average speeds.

Mode choice model estimation

To estimate a mode choice model, we used home-based trips to primary destinations, that is, work and place of education, from VISTA 2012-16 with origin and destination within the Greater Melbourne area as the main input. The reason for choosing home-based primary trips was to ensure that all modes of travel were available to travellers and to minimise the number of unaccounted factors in the choice model that affected the choice of the mode. Of these trips, those with driving, PT (including train, tram, and bus), walking or cycling as their mode of transport were selected. The centroids of the Australian Census Statistical

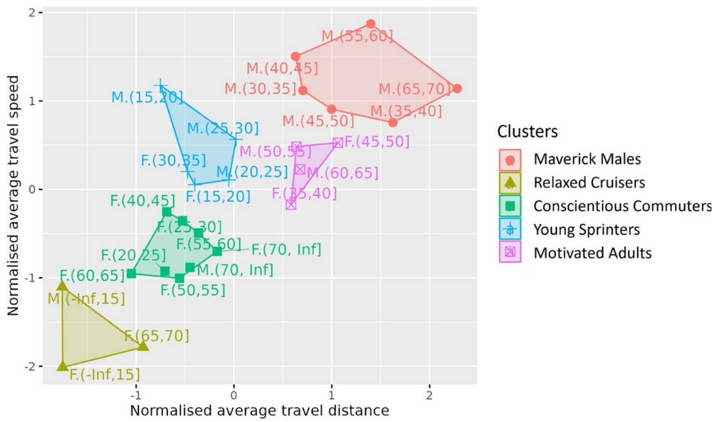


Fig. 6 Normalised cluster plot

Table 1 The demographic and behavioural profile of clusters

Label	Cohorts	N travellers all† (%)	N travellers cyclists‡ (%)	N trips§ (%)
Maverick Males	M.(30,50]*	7,457 (20.7)	300 (32.0)	25,971 (21.3)
	M.(55,60]			
	M.(65,70]			
Relaxed Cruisers	F.(65,70]	7,773 (21.6)	207 (22.1)	22,876 (18.8)
	F/M.[0,15]			
Conscientious Commuters	F.(20,30]	10,101 (28.0)	164 (17.5)	35,694 (29.3)
	F.(40,45]			
	F.(50,65]			
	F/M.(70,inf)			
Young Sprinters	F.(15,20]	5,341 (14.9)	132 (14.1)	16,545 (13.6)
	M.(15,30]			
	F.(30, 35]			
Motivated Adults	M.(50,55]	5,356 (14.9)	134 (14.3)	20,825 (17.1)
	M.(60,65]			
	F.(35,40]			
	F.(45,50]			
Total	-	36,028 (100)	937 (100)	121,911 (100)

* As an example, M.(35,40] refers to males aged higher than 35 and lower or equal to 40

†Persons recorded in VISTA 2012-16 with at least one trip with origin and destination within Greater Melbourne area

‡Persons recorded in VISTA 2012-16 with at least one bicycle trip with origin and destination within Greater Melbourne area

§Non-recreational trips recorded in VISTA 2012-16 with origin and destination with Greater Melbourne area

Area level 1 (SA1) of origin and destination for each of these trips were used as the origin/destination points of the trips. SA1 was the smallest spatial unit of reporting in the public version of the VISTA 2012-16 survey with an average radius of approximately 327 ms.

A flat fare was assumed for each PT trip, represented as Δm_{PT}. Also, for PT it was assumed that in addition to cost, travel time, $t_{trav,PT}$, is the main factor influencing mode choice (Eq. 4), rather than distance. For driving, we assumed that the effects of distance to be reflected through travel time, $t_{trav,Driving}$, and fuel consumption cost, $\Delta m_{Driving}$ (Eq. 3). The monetary cost of walking and cycling was considered zero. For walking, only travel time was included in the model as a variable. For cycling, the distance travelled on different types of cycleways was also incorporated into the mode choice function (Eq. 6).

Equation 3 outlines the utility functions of the mode choice model for different modes, where $\beta_{trav,i}$ denotes the marginal utility of travel time for mode *i*, β_m represents the marginal utility of money, $\beta_{inf(a)}$ is the marginal utility of travelling on a road segment *a* with cycling infrastructure type *inf*(*a*) and length l_a, and asc_i is the alternative-specific constant for mode *i*.

$$S_{trav,Driving} = \beta_{trav,Driving} \times t_{trav,Driving} + \beta_m \times \Delta m_{Driving}, \tag{3}$$

$$S_{trav,PT} = asc_{PT} + \beta_{trav,PT} \times t_{trav,PT} + \beta_m \times \Delta m_{PT}, \tag{4}$$

$$S_{trav,Walking} = asc_{Walking} + \beta_{trav,Walking} \times t_{trav,Walking}, \tag{5}$$

$$S_{trav,Cycling} = asc_{Cycling} + \beta_{trav,Cycling} \times t_{trav,Cycling} + \sum_{a \in A} (\beta_{inf(a)} \times l_a). \tag{6}$$

The approach proposed by Ziemke et al. (2019) was followed to include the impact of the type of cycling infrastructure in the choice model. Following this approach and as indicated in Eq. 7, for a road segment a, value of $\beta_{inf(a)}$ was determined based on a maximum marginal utility of the bicycle infrastructure, $\beta^{max}_{inf,Cycling}$ and an infrastructure factor $f_{inf}(a)$ representing the cycleway type that existed on the road segment a. The values for $f_{inf}(a)$ corresponding to different types of cycling infrastructure were directly obtained from Ziemke et al. (2019) and are detailed in Table 2. Separated and dedicated cycling infrastructures were assigned values closer to one, while main roads lacking such infrastructure received values near zero. Hence, according to Eqs. 7 and 6, the greater the separation of cycling infrastructure on road segment a, the lower the value of $\beta_{inf(a)}$, thereby reducing the infrastructure's impact on utility. In essence, the utility function captures the adverse effects of inadequate separation from motorised traffic on cycling, with the ideal scenario being no negative influence from the cycling infrastructure on the utility.

$$\beta_{inf(a)} = \beta^{max}_{inf,Cycling} \times (1 - f_{inf}(a)) \tag{7}$$

Travel time and cost for driving and public transport trips

The Distance Matrix API service[2] from the Google Maps platform was used to estimate the travel time and distance to drive and PT. This platform was selected because it incorporates the impact of congestion on trip projections and uses data from General Transit Feed Specification (GTFS) for the routing and timing of PT trips. It should be noted that the Distance Matrix API only accepts queries for a time in the present or the future. A midweek workday in November 2023 was selected to send queries. As such, there are potential inaccuracies introduced through the use of this API given the changes in the transport network since the travel survey was completed.

The cost of driving per kilometre for each trip was calculated based on the average fuel consumption of 11.8 L/100 km for a medium car according to the guidelines of The Australian Transport Assessment and Planning (ATAP) for the values of the road parameters,[3] and the average annual retail fuel price for 2016 in Victoria which was \$1.16 according to

Table 2 Bicycle infrastructure factor, $f_{inf}(a)$, for different road types from OpenStreetMap and their cycleway type based on (Ziemke et al. 2019)

Road type	Bicycle infrastructure factor $f_{inf}(a)$		
	Mixed	On-street lane	Bicycle path*
Trunk	0.05	0.95	-
Primary	0.1	0.95	-
Secondary	0.3	0.95	-
Tertiary	0.4	0.95	-
Unclassified	0.9	0.95	-
Cycleway	-	-	1.0

* "Bicycle path" is equivalent to the "Cycleway" road type in OSM

[2] https://developers.google.com/maps/documentation/distance-matrix/intro

[3] https://www.atap.gov.au/sites/default/files/pv2_road_parameter_values.pdf

data from the Australian Institute of Petroleum.[4] The PT travel cost was calculated under the assumption that each PT user has on average 2 PT trips per day. Therefore, given that the cost of the daily PT travel pass was \$7.80 in 2016, a travel cost of $7.80/2{=}3.90$ per trip was assumed for each PT trip.

Finding the route for cycling

We used the OSM-based road network discussed in Sect. 3 to route bicycle trips. We selected a connected subset of this network, consisting of roads where cycling was permitted (for instance, cycling is usually not allowed on footpaths and motorways in Melbourne), for routing.

Next, we used a heuristic route assignment approach to determine the best guess cycling route for each trip from VISTA. Subsequently, we identified the types of cycleways that constitute the assigned route for each trip, a necessary step to calculate the distance cycled on road segments with different types of bicycle infrastructure, as specified in Eq. 6. We used Strava Metro data (hereafter Metro data) for November 2023, which provided information on the weekday volume of bicycle commute traffic, to identify the best guess cycling route for each trip from VISTA. Each road segment with cycling volume data from Metro data was joined to the closest road segment in the transport network model.

A cycling popularity score for a road segment *a* with the recorded volume of cycling traffic of *n*, $P_a(n)$, was defined as a variable with values ranging from zero to one. We used a generalised symmetrical logistic function to calculate the value of $P_a(n)$ for each segment of the road. Assuming that the upper asymptote of y=1 and the lower asymptote of y=0, the logistic function presented in Eq. 8 was used to calculate the popularity score.

$$P_a(n) = \frac{1}{(1 + e^{-k(n-n_0)})}. \tag{8}$$

In Eq. 8, *k* determines the growth rate and n_0 denotes the sigmoid midpoint of the curve. The average number of bicycle riders per day for links with cycling allowed by Metro data was 21.4 and was used as n_0 for the logistic function.

Studies have shown that bicycle riders with varying levels of experience exhibit different preferences regarding the extent to which they are willing to lengthen their trips to use a road with better cycling infrastructure, such as off-road bicycle paths (Pucher et al. 2010; Reggiani et al. 2022; Rupi et al. 2019). Therefore, for each demographic cluster *i*, we defined a factor for the maximum deviation for bicycle riders to use the road segment with the highest popularity score, $f_{max,i}$. Let us assume that there are two routes from point A to B: one is the shortest route but with the lowest popularity, denoted R_{AB}, and the other is denoted as R'_{AB}, which is f_i times longer but has the highest popularity. For each cluster *i*, $f_{max,i}$ is the maximum value of f_i that bicycle riders in that cluster weakly prefer R'_{AB} over R_{AB}, that is, $R'_{AB} \succcurlyeq R_{AB}$.

Evidence from the literature supports that there are differences in the level deviation of the shortest route for cyclists of different demographic groups and experience levels(Buehler and Dill 2016). We assumed that demographic clusters with high cycling travel distance and speed are willing to deviate less from the shortest route. Therefore, the value of $f_{max,i}$, that

[4] https://aip.com.au/aip-annual-retail-price-data

is, the maximum theoretical deviation from the least popular shortest route to the most popular alternative route, was assumed to be 10% for Maverick Males, 20% for Young Sprinters, 30% for Motivated Adults, 40% for Conscientious Commuters, and 50% for Relaxed Cruisers.

It should be noted that $f_{max,i}$ indicates the theoretical maximum values for which bicycle riders in each cluster deviate from the shortest route in search of the most popular route, that is, the popularity score of $P_a(n) = 1$. As we will show later, this maximum deviation is a rare occurrence, and when these values are used for routing, the resulting deviations are much lower than the values specified earlier. We calculated the deviation factor $f_{a,i}$ for each road segment *a* with a popularity score of $P_a(n)$ and for bicycle riders in the demographic cluster of *i* as follows:

$$f_{a,i} = P_a(n) \times f_{max,i}. \tag{9}$$

We used $f_{a,i}$ to calculate the weighted length $l'_{a,i}$ for each road segment *a* with the actual length of l_a and for a cyclist in the demographic cluster *i* (Eq. 10).

$$l'_{a,i} = \frac{l_a}{(1 + f_{a,i})}. \tag{10}$$

We used an implementation of the Dijkstra algorithm in the R programming language to find the cycle route with the shortest weighted total length for each trip. The actual length of the routed road segments was used as the cycling distance for each trip. The travel time was calculated accordingly using the average cycling speed of each cluster.

The average resulted deviation for Maverick Males was 4.61% ($SD = 7.50\%$), for Young Sprinters it was 4.73% ($SD = 6.33\%$), for Motivated Adults it was 5.93% ($SD = 8.67\%$), for Conscientious Commuters it was 6.24% ($SD = 9.42\%$), and for Relaxed Cruisers it was 9.88% ($SD = 15.6\%$).

Bicycle traffic volume at a road segment level after the assignment of the cycling route was analysed to examine whether it represents the patterns observed in real-world data. To do this, we used automated cycling traffic volume count data from the Victorian Department of Transport. This data set includes hourly cycling traffic counts for 35 sites in Greater Melbourne. Data were filtered to records for weekday trips in November 2023 and used for further analysis. A Pearson correlation test was used to test whether there is no correlation between real-world cycling traffic volume data and the cycling volume resulting from the above-mentioned routing process. The correlation test showed that there is a significant association ($p < 0.001$) between the cycling volume values from the real-world data and the assigned cycling routes with the correlation coefficient value of 0.55, showing a positive correlation. For comparison, we also performed the routing using Dijkstra's shortest path algorithm and tested the correlation between the resulting cycling link-level traffic volume, which showed to have a weaker correlation than the routing based on our proposed approach, with a coefficient value of 0.36 ($p < 0.05$).

Finding the shortest path for walking

For walking, it was assumed that pedestrians use the shortest route to get to their destination. Furthermore, we assumed that traffic congestion, whether from car or pedestrian traffic, does not affect walking speed or route choice. A connected subset of the network explained in Sect. 4.2 was used for the routing, composed of only road segments where walking is permitted. We used Dijkstra's algorithm to find the shortest network path for each trip. The distance travelled for each trip was the total length of the road segments along the assigned shortest path, with travel time calculated assuming a constant walking speed of $1.7m/s$.

Coefficient estimation

Five separate mode choice models were estimated, one for each demographic group in which driving, PT, walking, and cycling were considered as alternatives. Travellers without a car or bicycle in their household had these travel mode options removed from their choice set. A maximum logarithmic likelihood Multinomial Logit model (MNL) model was used to estimate the choice model coefficients for each cluster (Molloy et al. 2019).

Table 3 provides a summary of the results of the choice model for each cluster. The estimated coefficients of the mode choice model in Table 3 indicated that travel time is a significant factor in the selection of the mode of travel for the five clusters. However, cycling infrastructure was not found to be significant for the Motivated Adults cluster. Similarly, cost was found to be non-significant for travellers from the Relaxed Cruisers cluster, which consists of travellers under 15 years old as well as females aged 65 to 70. These coefficients were used to build the choice model for the simulation model, as described in the next section.

Negative constant values ($asc_{Cycling}$) for all groups except relaxed cruisers indicate that when all other variables in the model are set to zero, the baseline utility of choosing cycling over the reference alternative is negative for those groups. The negative sign of the coefficients $\beta^{max}_{inf,Cycling}$ indicates a reduction in utility given the cycling infrastructure, and the

Table 3 Estimated mode choice model coefficients

	Estimation (Standard error)				
Coefficients	Maverick Males (N=3415)	Relaxed Cruisers (N=767)	Conscientious Commuters (N=3076)	Young Sprinters (N=2176)	Motivated Adults (N=2469)
β_m	-0.35^{***} (0.07)	0.14 (0.25)	-0.52^{***} (0.08)	-0.19^{**} (0.07)	-0.17^{*}(0.10)
asc_{PT}	-0.78^{**} (0.28)	0.49 (1.17)	-0.15 (0.38)	-0.53 (0.36)	-2.31^{***}(0.48)
$asc_{Walking}$	-0.13 (0.35)	6.45^{***} (0.43)	0.57^{**} (0.26)	0.98^{***} (0.23)	0.33(0.32)
$asc_{Cycling}$	-1.92^{***} (0.17)	3.66^{***} (0.39)	-2.54^{***} (0.26)	-1.31^{***} (0.23)	-3.27^{***}(0.26)
$\beta_{trav,Driving}$	-4.04^{***} (0.25)	-3.10^{***} (0.99)	-4.39^{***} (0.30)	-3.41^{***} (0.28)	-4.88^{***}(0.42)
$\beta_{trav,PT}$	-5.58^{***} (0.43)	-3.67^{***} (1.20)	-6.30^{***} (0.55)	-4.15^{***} (0.46)	-6.82^{***}(0.86)
$\beta_{trav,Walking}$	-6.61^{***} (0.67)	-10.45^{***} (0.95)	-8.14^{***} (0.82)	-6.64^{***} (0.56)	-8.58^{***}(1.11)
$\beta_{trav,Cycling}$	-4.54^{***} (0.48)	-7.06^{***} (1.32)	-3.67^{***} (0.61)	-3.91^{***} (0.59)	-3.83^{***}(0.63)
$\beta^{max}_{inf,Cycling}$	-0.11^{*} (0.07)	-0.38^{*} (0.37)	-0.49^{**} (0.20)	-0.17^{*} (0.11)	-0.12(0.11)
R2	0.63	0.54	0.63	0.46	0.72

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

magnitude of the coefficient indicates the extent of this reduction. For example, based on the bicycle infrastructure factors presented in Table 3, for the Maverick group, when using mixed traffic trunk roads for cycling, a value of $-0.11 \times (1 - 0.05) = -0.1045$ times the length of the road segment would be deducted from the utility for cycling, and for tertiary roads, a value of $-0.11 \times (1 - 0.4) = -0.066$ times the length of the road segment would be deducted. Therefore, the best-case scenario would be when no negative value is added to the utility due to the cycling infrastructure, that is, for a cycleway with $f_{inf}(a) = 1$.

Modelling the strategic cycling corridors scenario

Strategic Cycling Corridors (SCC) are a set of existing and planned cycling corridors across greater Melbourne that provide a connection between major destinations, including Melbourne CBD, employment hubs, and activity centres (Department of Transport 2020). They were prioritised in key cycling-related strategic documents for Melbourne including Plan Melbourne 2017-2050 (Victoria State Government 2017) and Victorian Cycling Strategy 2018-28 (Department of Transport 2018), highlighting the importance of understanding the potential impact of these corridors. Some of the road segments in SCC already have a cycleway and are set to be upgraded, while others are planned for new cycleways. The main goal of SCC is to fill the network gaps by adding new infrastructure where cycleways are missing and upgrading existing ones if necessary to create a safe and connected network for bicycle riders.

Mapping the strategic cycling corridors

The 2020 version of SCC was obtained from the Victorian Department of Transport (DoT) website[5] and was used as input for designing the intervention scenario. The obtained map layer consists of simplified line geometries that only approximately represent the roads of the real-world transport network. Therefore, it was necessary to devise a process to spatially join the SCC with the OSM based transport network in order to fully integrate them.

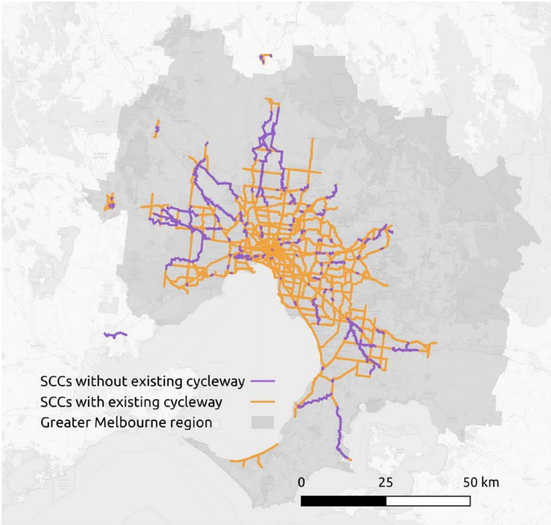
To perform the spatial join, a buffer with a radius of 250 m was initially created for each corridor in the SCC map layer. Next, to reduce the search space, the road segments from the transport network were filtered to only those that intersected their buffer. Start and end points of the corridors were then snapped to the closest point on the filtered transport network and the shortest path was calculated between them. If a route was found, all road segments of the transport network along that route were marked as being part of the SCC. An on-street bicycle lane was added to these road segments if they lacked any type of cycleway. If they already had an on-street bicycle lane or a dedicated bike path, then the cycleway type attribute was not changed. If no route was found, it was assumed that the corridor (or part of it) represented a dedicated off-road bicycle path that has yet to be built. In this case, Dijkstra's shortest path algorithm was used to find suitable paths from both start and end points, attempting to cover as much of the corridor as possible using the existing road network until a gap was encountered. The gap was then bridged by creating new links that represent the missing section of the new segment as a new off-road cycle path. To ensure connectivity with the simulation model network, every 250 ms along these new off-road cycle paths, a

[5] https://www.vic.gov.au/strategic-cycling-corridors, Accessed: 28/02/2024.

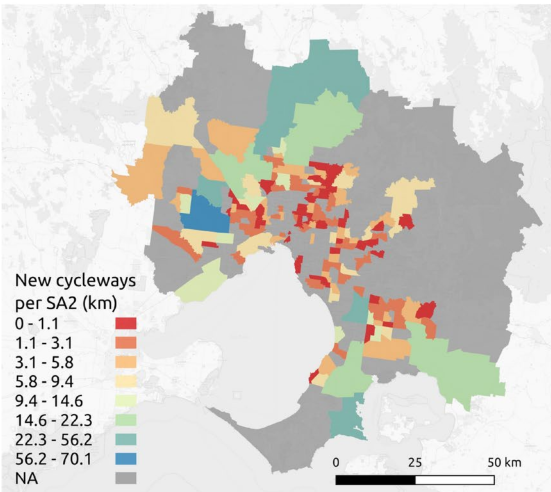
connection point was created to the nearest existing road network, so cyclists could enter and exit this new off-road path.

Figure 7(a) depicts the SCC joined to the simulation model's transport network. Additionally, Fig. 7(b) shows the spatial distribution of the new cycleways that our analysis indicated will be added, aggregated to the SA2 level. This figure illustrates that most of the new cycling infrastructure will be in the middle and outer ring areas of Greater Melbourne, which are often known as areas significantly lacking good cycling infrastructure. However, these areas are also characterised by having lower residential density and longer trip distances, which make them less attractive for cycling compared to the inner ring suburbs of Melbourne.

Fig. 7 Distribution of the strategic cycling corridors across Greater Melbourne region based (basemap from OpenStreetMap)



(a) SCC network joined



(b) New cycleways per SA2

Analysing the change in access to cycling infrastructure

To better understand the interrelationships between the existing infrastructure and the support for current trips within a bikeable distance, we routed trips less than 5 km from the VISTA on a subset of the road network where cycling is allowed. Trips in VISTA are aggregated and reported at the SA1 level to protect the privacy of the respondents. We used centroids of origin and destination SA1s as start and end points for each trip. The shortest route on the sub-network was calculated for each origin–destination pair using Dijkstra's shortest path algorithm.

Figure 8 shows the average percentage of the length of routes that were routed on roads with any type of cycleway, aggregated based on the trip origin SA2, before and after the full implementation of the SCC scenario. As illustrated in this figure, and when compared with Fig. 7(b), it is evident that in areas where new cycleways are expected to be built, an improvement in the coverage of the shortest routes for trips less than 5 km with cycling infrastructure is predicted. These suburbs experiencing improvements are primarily middle-ring suburbs and those around main population centres in the outer-ring suburbs of Melbourne.

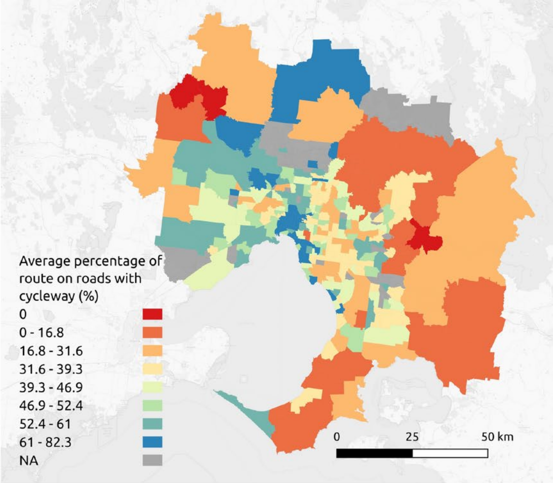
Additionally, to understand these changes in route coverage before and after the full implementation of the SCC scenario, we examined the average length of trips under 5 km in Greater Melbourne that were routed on roads with any type of cycleway, as well as the percentage of the route for different demographic clusters. As shown in Table 4, this scenario resulted in an approximate 30% increase in the cycleway coverage percentage of the routed trips for almost all clusters. However, even after this increase, only just over half of the lengths of routes are on average on roads with any cycleway, ranging from 50.9% for Relaxed Cruisers to 54.8% for Young Sprinters.

Estimating the mode shift using a simulation model

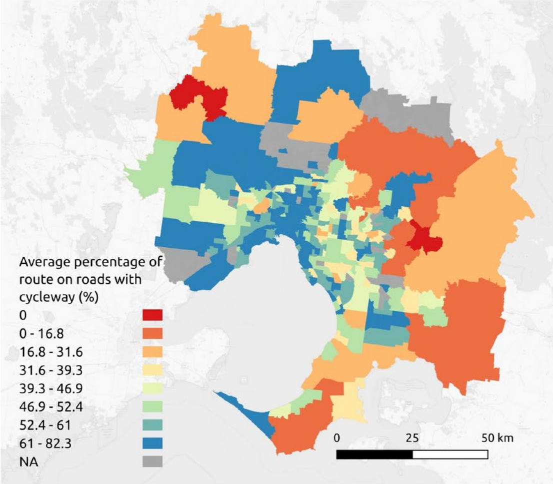
Next, we examined the expected increase in cycling mode share as a result of the increase in cycleway coverage in the SCC scenario using a simulation model. Although the mode choice described in Sect. 4 can be useful in estimating the mode shift, it does not consider the dynamics of the transport system, such as changes in congestion that would likely result from the shift towards cycling. Therefore, we used an agent- and activity-based transport model to investigate the mode change resulting from the SCC scenario. Furthermore, the simulation model, with its synthetic population of travelling agents, enables an investigation into the impact of a much larger population than the relatively small sample size of VISTA.

The open-source model development workflow proposed by Jafari et al. (2024) was followed to build the open scenario of the Activity-based and agent-based Transport model **o**f Melbourne (AToM). The AToM open scenario includes a calibrated 10% mid-week day activity-based transport demand for Melbourne, which is explained in detail by Both et al. (2021). The resulting transport demand is a list of virtual travellers with their age, sex, household composition, and employment status (employed or unemployed), as well as their travel diaries, including a chain of activities for the day, the location and timing of each activity, and the travel mode for reaching each activity. Both et al. (2021) provides a detailed description of the calibration and validation process and its results, showing a good match to real-world observation data in terms of travel distances, activity locations and timing,

Fig. 8 Average percentage of length of the shortest routes for trips less than 5 km that are on roads with any type of cycleway aggregated based on trip origin SA1 for **a** full implementation of the SCC and **b** after full implementation of the SCC (basemap from OpenStreetMap)



(a) Before fully implemented SCC



(b) After fully implemented SCC

Table 4 Average length and percentage of the shortest route of bikeable trips (< 5km) that are routed on roads with any type of cycleway for difference clusters before and after the full implementation of the SCCs

	Length (km)			Percentage (%)		
Cluster	Before	After	Change	Before	After	Change
Maverick Males	1.33	1.72	+27.0%	42.1	53.4	+29.6%
Relaxed Cruisers	1.21	1.57	+25.1%	40.7	50.9	+29.5%
Conscientious Commuters	1.31	1.75	+28.5%	41.3	53.1	+33.5%
Young Sprinters	1.44	1.93	+28.9%	42.5	54.8	+33.8%
Motivated Adults	1.32	1.74	+27.6%	40.6	51.8	+31.4%
Total	1.31	1.72	+27.3%	41.3	52.6	+31.6%

as well as demographic features. Using this activity-based transport demand, we assigned each virtual traveller (agent) to one of the clusters presented in Table 1 based on their sex and age cohorts.

Another tool presented in this AToM open scenario development workflow was an open-source transport network generation tool for large-scale active transport simulations, which is explained in detail by Jafari et al. (2022). This is the tool that we used previously to build a road network for routing the mode choice model described in Sect. 4. This tool was also used to build a road network model for Greater Melbourne from the OSM extract of the area for the simulation model. The output network included typical road attributes necessary for a Multi-Agent Transport Simulation (MATSim) model, such as speed limits, road capacity, number of lanes, and cycleway types. Additionally, data from GTFS were used to add PT stops and routes to the transport network, as well as to extract schedules for each PT service to be used in the simulation model.

A simple mode choice model for the four main modes of car, PT, walking, and cycling with only parameters for travel time and cost was the third input included in the AToM open scenario development workflow. Rather than using the simple mode choice model introduced by Jafari et al. (2024), we used the estimated coefficients of the clustered mode-choice model presented in Table 3 for the simulation model's parameters. This enables us to test the impact of bicycle infrastructure change scenarios, such as SCCs, on different demographic clusters using the simulation model.

Similar to the approach suggested by Jafari et al. (2024), MATSim (version 13.0) was used to simulate the transport system. The bicycle extension for MATSim, developed by Ziemke et al. (2019), was used to incorporate the bicycle infrastructure coefficient in the utility function of the MATSim (Eq. 6). Cycling and driving movements were explicitly simulated on the network. However, the direct negative impact of interactions with motorised traffic on cycling utility was not considered, which is a limitation of this study. PT travel was set to follow a deterministic schedule extracted from GTFS, independent of other modes of travel. Walking was treated as a teleportation mode, meaning that walking agents disregarded the road network and used the bee-line distance from origin to destination. A distance correction factor of 1.3 was applied to walking to account for the difference between the network distance and the simulated teleportation distance, based on an analysis of walking trips in VISTA from 2012-16, which showed that the recorded network distance was, on average, 30% longer than the bee-line distance between the centroids of the origin/destination SA1 regions.

The simulation model was initially run for 200 iterations, followed by 10 runs of 100 iterations for model calibration. Two innovation or mutation strategies, rerouting and sub-tour mode choice, were enabled for the first 80% of iterations. The occurrence chance for each of these innovation strategies was set to 10%. This meant that during the first 80% of iterations, after each iteration i, there was a 10% chance that agent j would try an alternative route for trip k, and a 10% chance of using a different mode of travel than previously selected. MATSim's randomising router was used to explore alternative routes for each agent in different iterations of the simulation. Driving and cycling were set as *tour modes*, meaning that if agents left the house using either of these two modes, they had to return using the same mode. At the end of each iteration, all the executed travel plans were scored based on the utility function coefficients. No new plans (i.e., new routes or travel modes)

Table 5 Mode share comparison before and after of the SCC scenario for different demographic clusters

		Mode share (%)			
Cluster	Travel Mode	VISTA	Baseline	Scenario	Change (%)
Maverick Males	Car	77.2	74.1	73.3	−1.06
	Public Transport	7.2	8.0	7.8	−2.45
	Bicycle	2.2	2.3	3.3	+42.62
	Walk	13.4	15.6	15.6	+0.00
Relaxed Cruisers	Car	79.1	76.1	75.7	−0.53
	Public Transport	3.6	4.5	4.4	−2.2
	Bicycle	1.5	1.8	2.4	+32.2
	Walk	15.8	17.7	17.5	−0.98
Conscientious Commuters	Car	77.2	73.9	72.2	−2.35
	Public Transport	6.4	7.9	8.4	+6.33
	Bicycle	1.0	1.2	1.8	+52.95
	Walk	15.4	17.1	17.6	+2.92
Young Sprinters	Car	70.8	65.9	65.7	−0.34
	Public Transport	12.8	14.8	14.6	−1.15
	Bicycle	1.7	2.5	3.1	+22.69
	Walk	14.7	16.8	16.6	−1.01
Motivated Adults	Car	80.1	74.7	74.6	−0.13
	Public Transport	4.7	5.9	6.00	+0.95
	Bicycle	1.7	3.0	3.10	+3.60
	Walk	14.0	16.4	16.30	−0.72
Total	Car	76.9	73.3	72.5	−1.02
	Public Transport	6.9	7.9	8.0	+0.95
	Bicycle	1.6	2.0	2.6	+30.64
	Walk	14.6	16.8	16.9	+0.39

were allowed for the last 20% of iterations. The remaining configurations of the simulation model were similar to those used by Jafari et al. (2024).

Given that the new and clustered mode choice parameters were the main difference between the simulation model used in this study and the AToM baseline scenario, we used VISTA 2016-20 data to analyse and re-calibrate the mode choice behaviour of the simulation model to match the observations from the real world. A detailed discussion on the calibration of other components of the AToM open scenario is presented in Jafari et al. (2024). Trips that started and ended within the Greater Melbourne area and were carried out using one of the four modes of transport-driving, PT, walking, and cycling-were included. These trips were grouped into five clusters based on the demographic characteristics of the travellers, and the share of transport modes was calculated for each cluster. The resulting mode share for each cluster from VISTA was compared with the output of the simulation model. The alternative-specific constants for each mode were adjusted to calibrate the mode choice behaviour of the model through an iterative process until the mode shares from the simulation outputs reasonably matched those from VISTA. Table 5 presents the resulting mode share percentages of the calibrated simulation model compared with VISTA 2016-20 data for different demographic clusters, showing that across all clusters, the simulation model resembles the mode share distribution patterns observed in the real-world data.

The calibrated simulation model was then used to examine the impact of full SCC implementation in Greater Melbourne (see Fig. 7(a)) for each of the demographic clusters, with

Table 5 providing an overview of their change in cycling mode share. As shown in the table, the full implementation of the SCC resulted in a 30.64% increase in cycling mode share for the total population across Greater Melbourne. The highest increase in cycling mode share was observed among the Conscientious Commuters and Maverick Males, with increases of 52.95% and 42.62%, respectively. These clusters consist mainly of middle-aged female and male travellers, indicating the significant potential of the SCC to encourage people from these demographic groups to switch to cycling.

The smallest increase was observed in the Motivated Adults cluster, which was expected since the impact of cycling infrastructure on mode choice for this cluster was not found to be statistically significant in the mode choice model (Table 3). Among the clusters where cycling infrastructure was a significant attribute in mode choice, Young Sprinters had the lowest simulated mode shift to cycling, with a 22.69% increase.

Notably, increases in cycling across all clusters resulted in a reduction in car usage, with a driving mode share change of −1.02% for all clusters combined. Public transport use also increased by 0.95%, and walking mode share increased slightly, with a total change of 0.39%. This indicates that a more diverse set of travel modes can be used to reach daily destinations once private vehicles drop out of a trip chain.

Discussion and conclusion

Demographic attributes, particularly age and sex, have been widely considered in studies using multivariate statistical methods to investigate cyclist behaviour (Branion-Calles et al. 2019; Nehme et al. 2016; Zahran et al. 2008; Zhao 2014), and have been shown to be significant determinants in choosing to cycle. However, these attributes have rarely been included in city-scale transport models, primarily due to the lack of individual-level cycling route data in most cities worldwide and the complexity of capturing heterogeneous decision-making in a simulation model. Consequently, the use of transport models to study potential changes in cycling adoption due to changes in infrastructure among different demographic groups has until now been limited.

In this study, a cluster-based approach was proposed to incorporate differences between demographic cohorts in mode choice modelling and a city-scale agent-based model. Cycling travel distance and speed, along with the traveller's age and sex, were the variables included in the clustering. These variables can commonly be found in travel surveys of cities worldwide. Therefore, although the clustering was performed for Melbourne, Australia, the same process can be followed to cluster bicycle riders in other cities globally.

In total, five clusters of bicycle riders were identified for Melbourne. Other studies have also explored ways to cluster bicycle riders, with one of the most well-known studies finding four clusters of bicycle riders proposed by Dill and McNeil (2013), which include No Way No How, Interested but Concerned, Enthused and Confident, and Strong and Fearless. More recently, Fraboni et al. (2022) used cluster analysis to categorise bicycle riders according to their frequency, purpose, and attitudes toward cycling, along with other demographic and environmental factors. A key difference between the approach of the current paper and those previously is that the purpose of the clustering analysis in this paper was specifically geared toward building a choice model rather than identifying differences in cycling trip behaviour. Therefore, the included variables were also the key variables in the choice model.

The lack of route data for individual bicycle riders has been another major obstacle in building city-scale cycling models as it limits the possibility of including route-based attributes in the utility function of the model. Having this information is necessary to assign a cycling route to each trip, as evidence from the literature supports that bicycle riders often cycle longer distances to use better and safer infrastructure (Pucher et al. 2010). To address this, we used cycling traffic volume data from Strava Metro and assigned a best guess route to each of the observations in the travel survey data. Compared with real-world cycling volume sensor data, we showed that our approach to calculate the best guess cycling route more accurately represents real-world cycling volumes than simple shortest path routing for cycling. For car driving and public transport, we used the Google Maps API, and for walking, we used the shortest network path. Given that all data sets used for this routing process are universal, the process proposed in this study can be adopted in other cities around the world to assign best guess routes to origin–destination pairs from travel surveys or other sources.

The MNL mode choice model using the routed travel survey data showed that for four of the five demographic clusters, cycling infrastructure is a statistically significant factor ($p < 0.1$) in choosing which mode to use. The only cluster where bicycle infrastructure was not found to be a significant factor was the Motivated Adults cluster, which consists of males aged 50-55 and 60-65, and females aged 35-40 and 60-65. This cluster represents the cyclists who travel above average distances, and are of working age, but mostly in the upper range of working age. In the travel survey, this cluster had only 134 cyclists. The low number of cyclists in the sample size for this cluster could be one of the reasons that the statistical test did not show a significant result. The result of the choice model shows that, in addition to typical coefficients for travel time ($p < 0.001$) and cost ($p < 0.1$) that were found to be significant factors associated with mode choice across all clusters (except travel cost for the Relaxed Cruisers cluster), cycling infrastructure is another key coefficient whose impact needs to be considered.

Having different demographic clusters for the mode choice model rather than including the demographic attributes in the choice model itself as an independent variable makes it more straightforward to integrate the estimated coefficients into an agent-based transport simulation model, such as a MATSim model. Using agent-based simulation models built on the estimated mode choice model makes it possible to include the impact of changes in the transport system, e.g., traffic congestion and subsequently travel time, as a result of a change in mode share when looking at the impact of different scenarios on mode share. To illustrate this, in this study, we used the estimated mode choice parameters in a city-scale agent- and activity-based transport simulation model and showed the usefulness of the model by using it to test the impact of new cycling infrastructure in Greater Melbourne on mode share.

We used the full implementation of SCCs in Greater Melbourne as our built-environmental change scenario and tested its impact on mode change. The before and after comparison of the simulation results showed that an uptake of, on average, 30.64% in cycling mode share can be expected as a result of the full implementation of the SCCs across all demographic groups, with the highest expected increase among Conscientious Commuters with about a 52.95% expected increase in cycling mode share. Although these numbers show an increase in cycling use, the cycling mode share still remains very low at 2.6%, which is considered similar to the cycling rates of low cycling countries (Buehler and Pucher 2021).

To better understand the reason for this modest increase, despite the scale of the SCC, we looked at all trips with a distance of 5 km or less in the travel survey data and routed them on the road network before and after the SCC scenario. Comparing the routed trips, we found that although SCC will result in, on average, a 31.6% increase in the cycleway route coverage percentage of the shortest route for these trips to be on roads with a cycleway, still about 47.4% of the length of these routes will be on roads without any type of cycleway. This shows that although they provide a connected network of cycleways around the main population centres of Greater Melbourne, they do not address the continuity gap of the cycling infrastructure for travellers throughout the entire trip, which is a factor that has been found to be critical for cycling (Heinen et al. 2010). Therefore, more research is needed to understand how a feeder network of cycleways can complement SCC to provide adequate coverage of cycling infrastructure throughout the trip.

The findings highlight two key policy implications for improving cycling infrastructure in Melbourne. The first relates to existing urban infrastructure governance challenges in Australian cities and the lack of an integrated governance model across different levels of government (Clements et al. 2023; Steele 2020). The SCCs are primarily on arterial roads managed by the Victorian State's Department of Transport, while the local roads feeding into these corridors are managed by local councils. This division of responsibilities necessitates coordinated action between local and state governments to create a connected and safe cycling environment. Second, while this study focused on bike lanes, other measures could enhance safety, particularly on feeder networks connecting to the main corridors. Solutions such as lowering speed limits on residential streets and implementing traffic calming measures-such as the 30 km/h speed limit trial in the City of Yarra, one of the local government areas in the Greater Melbourne region (Lawrence et al. 2020)-would help address the infrastructure gaps identified in this study and support increased cycling adoption across Melbourne.

The current study has a number of limitations that should be considered. First, the data sets used in this study each had several limitations. This includes a low number of cycling trips in the travel survey data, particularly for some demographic clusters, as discussed earlier. This is a common limitation among cities with low cycling share such as Melbourne. The Victorian government has acknowledged this limitation and has better data on cycling as one of its key strategic actions (Department of Transport 2018). Additionally, Strava Metro data, which we used as a source for cycling traffic volume and are also used by others (Lee and Sener 2021), has limitations. A recent analysis of Strava Metro data in Oslo, Norway, showed that it can accurately capture spatial and temporal variation in recreational activities, but it under-represents the groups of youth, elderly, and those of low socioeconomic status (Venter et al. 2023). However, it is expected that this type of crowdsourced smartphone application data will become increasingly widely used by urban and transport planners in areas such as cycling, the lack of which has been a major obstacle in planning and research (Lee and Sener 2021; Venter et al. 2023). Due to the lack of route-level real-world data, we made several assumptions in this study, such as the deviation factors for the cycling routing component, which although overall supported by the literature, these assumptions are not context-specific. This indicates the need for granular route-level cycling data to support data-driven modelling for cycling.

Furthermore, the reliance on smartphone application data in this study may introduce biases, as it might not fully represent all demographic groups, particularly those less techsavvy or underrepresented. Future studies are encouraged to consider alternative data sources, such as GPS-enabled surveys or custom travel diaries, to improve the accuracy and representativeness of cycling behaviour analysis. Potential methods for collecting this data could include using GPS devices attached to bicycles (Gadsby and Watkins 2020), or wearable GPS and sensors for cyclists (Chou et al. 2023), with a representative sample recruited by the researchers. Additionally, combining these methods with other data sources, such as GPS recordings from bike share systems (Scott et al. 2021), could provide more precise and inclusive insights into cycling patterns across diverse populations and bicycle user types.

Another key limitation of this study is the limited number of demographic and road attributes considered in the clustering and mode choice modelling, and subsequently in the simulation model. Numerous studies have shown that a wide range of social and environmental factors influence cycling behaviour beyond age, sex, and the presence of bicycle-specific infrastructure. These factors include other demographic variables such as income level, occupation type, and fitness level, as well as additional built-environment factors beyond just the existence of bike lanes, such as the quality of bicycle lanes and end-of-trip facilities. Natural environment factors like hilliness of the road and weather conditions, along with perceptions, attitudes, and social influences, also play significant roles in cycling behaviour (Handy et al. 2010; Sallis et al. 2013; Willis et al. 2015; Boulange et al. 2017).

While it might seem advantageous to include a broader set of variables, there is a necessary trade-off in modelling between the level of detail and the model's practicality and interpretability (Sun et al. 2016). Including additional factors could improve the accuracy of the model but would also increase the complexity and data requirements for model development, and limit the interpretability of the model outputs due to this added complexity. This study, therefore, focused on key attributes that are consistently associated with cycling mode choice, balancing complexity with usability. Consequently, the simulation outcomes should not be viewed as precise forecasts of future cycling demand but rather as tools for gaining insights into cycling dynamics, supporting better decision-making, and enhancing the understanding of this complex phenomenon. Future research should aim to investigate a broader array of socio-demographic, psychological, and environmental variables-such as income, education level, and social norms-to enhance the model's ability to reflect the complex and evolving nature of cycling behaviour across diverse populations. None of these were included in this study. Additionally, our results primarily capture the initial response to infrastructure investments and do not account for evolving social norms and other behavioural changes over time that may encourage more people to cycle, a dynamic shown to significantly impact cycling adoption, which is another potential future avenue of research.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions

References

Aldred, R., Elliott, B., Woodcock, J., Goodman, A.: Cycling provision separated from motor traffic: a systematic review exploring whether stated preferences vary by gender and age. Transp. Rev. **37**(1), 29–55 (2017)

Aziz, H.A., Park, B.H., Morton, A., Stewart, R.N., Hilliard, M., Maness, M.: A high resolution agent-based model to support walk-bicycle infrastructure investment decisions: A case study with new york city. Transp. Res. Part C: Emerg. Technol. **86**, 280–299 (2018)

Both, A., Gunn, L., Higgs, C., Davern, M., Jafari, A., Boulange, C., Giles-Corti, B.: Achieving 'active'30 minute cities: How feasible is it to reach work within 30 minutes using active transport modes? ISPRS Int. J. Geo Inf. **11**(1), 58 (2022)

Both, A., Singh, D., Jafari, A., Giles-Corti, B., and Gunn, L. (2021). An activity-based model of transport demand for greater melbourne. arXiv preprint arXiv:2111.10061

Boulange, C., Gunn, L., Giles-Corti, B., Mavoa, S., Pettit, C., Badland, H.: Examining associations between urban design attributes and transport mode choice for walking, cycling, public transport and private motor vehicle trips. J. Transp. Health **6**, 155–166 (2017)

Branion-Calles, M., Nelson, T., Fuller, D., Gauvin, L., Winters, M.: Associations between individual characteristics, availability of bicycle infrastructure, and city-wide safety perceptions of bicycling: A cross-sectional survey of bicyclists in 6 canadian and us cities. Transp. Res. Part A: Policy Practice **123**, 229–239 (2019)

Braun, L.M., Rodriguez, D.A., Cole-Hunter, T., Ambros, A., Donaire-Gonzalez, D., Jerrett, M., Mendez, M.A., Nieuwenhuijsen, M.J., de Nazelle, A.: Short-term planning and policy interventions to promote cycling in urban centers: Findings from a commute mode choice analysis in barcelona, spain. Transp. Res. Part A: Policy Practice **89**, 164–183 (2016)

Buehler, R., Dill, J.: Bikeway networks: A review of effects on cycling. Transp. Rev. **36**(1), 9–27 (2016)

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

Buehler, R. and Pucher, J. (2021a). 21 cycling to a more sustainable transport future. *Cycling for Sustainable Cities*, page 101

Buehler, R. and Pucher, J. (2021b). International overview of cycling. *Cycling for Sustainable Cities*, pages 11–34

Chou, K.-Y., Paulsen, M., Nielsen, O.A., Jensen, A.F.: Analysis of cycling accessibility using detour ratios-a large-scale study based on crowdsourced gps data. Sustain. Cities Soc. **93**, 104500 (2023)

Clements, R., Alizadeh, T., Kamruzzaman, L., Searle, G., Legacy, C.: A systematic literature review of infrastructure governance: cross-sectoral lessons for transformative governance approaches. J. Plan. Lit. **38**(1), 70–87 (2023)

Crosland, P., Ananthapavan, J., Davison, J., Lambert, M., Carter, R.: The economic cost of preventable disease in australia: a systematic review of estimates and methods. Aust. N. Z. J. Public Health **43**(5), 484–495 (2019)

Debnath, A.K., Haworth, N., Heesch, K.C.: Women cycling in queensland: Results from an observational study. Acc. Anal. Prevent. **151**, 105980 (2021)

Department of Transport: Victorian Cycling Strategy 2018–28. Technical report, Departmnet of Transport (2018)

Department of Transport: Strategic cycling corridor network overview. Technical report, Departmnet of Transport (2020)

Dill, J., McNeil, N.: Four types of cyclists? examination of typology for better understanding of bicycling behavior and potential. Transp. Res. Rec. **2387**(1), 129–138 (2013)

Félix, R., Cambra, P., Moura, F.: Build it and give 'em bikes, and they will come: The effects of cycling infrastructure and bike-sharing system in lisbon. Case Studies Transp. Policy **8**(2), 672–682 (2020)

Fraboni, F., Prati, G., Casu, G., De Angelis, M., Pietrantoni, L.: A cluster analysis of cyclists in europe: common patterns, behaviours, and attitudes. Transportation **49**(2), 591–620 (2022)

Furth, P. G. (2021). Bicycling infrastructure for all. *Cycling for Sustainable Cities; Buehler, R., Pucher, J., Eds*

Furth, P.G., Mekuria, M.C., Nixon, H.: Network connectivity for low-stress bicycling. Transp. Res. Rec. **2587**(1), 41–49 (2016)

Gadsby, A., Watkins, K.: Instrumented bikes and their use in studies on transportation behaviour, safety, and maintenance. Transp. Rev. **40**(6), 774–795 (2020)

Garrard, J. (2021). 11 women and cycling: Addressing the gender gap. *Cycling for Sustainable Cities*

Garrard, J., Rissel, C., Bauman, A.: Health benefits cycling. City Cycling **31**, 31–56 (2012)

Giles-Corti, B., Saghapour, T., Turrell, G., Gunn, L., Both, A., Lowe, M., Rozek, J., Roberts, R., Hooper, P., Butt, A., et al.: Spatial and socioeconomic inequities in liveability in australia's 21 largest cities: Does city size matter? Health Place **78**, 102899 (2022)

Goel, R., Goodman, A., Aldred, R., Nakamura, R., Tatah, L., Garcia, L.M.T., Zapata-Diomedi, B., de Sa, T.H., Tiwari, G., de Nazelle, A., et al.: Cycling behaviour in 17 countries across 6 continents: levels of cycling, who cycles, for what purpose, and how far? Transp. Rev. **42**(1), 58–81 (2022)

Handy, S.L., Xing, Y., Buehler, T.J.: Factors associated with bicycle ownership and use: a study of six small us cities. Transportation **37**(6), 967–985 (2010)

Heesch, K.C., Giles-Corti, B., Turrell, G.: Cycling for transport and recreation: associations with socio-economic position, environmental perceptions, and psychological disposition. Prev. Med. **63**, 29–35 (2014)

Heesch, K.C., Sahlqvist, S., Garrard, J.: Gender differences in recreational and transport cycling: a cross-sectional mixed-methods comparison of cycling patterns, motivators, and constraints. Int. J. Behav. Nutr. Phys. Act. **9**(1), 1–12 (2012)

Heesch, K.C., Turrell, G.: Key influences on cycling for transport. J. Australasian College Road Safe. **25**, 4 (2014)

Heinen, E., Van Wee, B., Maat, K.: Commuting by bicycle: an overview of the literature. Transp. Rev. **30**(1), 59–96 (2010)

Hollauer, C., Lang, C., Wilberg, J., Weking, J., Dengler, C., Böhm, M., Krcmar, H., Lohmann, B., Omer, M., et al. (2018). Data-based development of an agent-based simulation to support the design of bicycle-sharing systems. In *DS 96: The 20th International DSM Conference*, pages 117–127

Jafari, A., Both, A., Singh, D., Gunn, L., Giles-Corti, B.: Building the road network for city-scale active transport simulation models. Simul. Model. Pract. Theory **114**, 102398 (2022)

Jafari, A., Singh, D., Both, A., Abdollahyar, M., Gunn, L., Pemberton, S., and Giles-Corti, B. (2024). Activity-based and agent-based transport model of melbourne: an open multi-modal transport simulation model for greater melbourne. *Journal of Intelligent Transportation Systems*, pages 1–18

Kraus, S., Koch, N.: Provisional covid-19 infrastructure induces large, rapid increases in cycling. Proc. Natl. Acad. Sci. **118**(15), e2024399118 (2021)

Lawrence, B., Fildes, B., Thompson, L., Cook, J., Newstead, S.: Evaluation of the 30km/h speed limit trial in the city of yarra, melbourne, australia. Traffic Inj. Prev. **21**(sup1), S96–S101 (2020)

Lee, K., Sener, I.N.: Strava metro data for bicycle monitoring: a literature review. Transp. Rev. **41**(1), 27–47 (2021)

Liu, C., Tapani, A., Kristoffersson, I., Rydergren, C., Jonsson, D.: Development of a large-scale transport model with focus on cycling. Transp. Res. Part A: Policy Pract. **134**, 164–183 (2020)

Llorca, C., Moeckel, R.: Assesment of the potential of cargo bikes and electrification for last-mile parcel delivery by means of simulation of urban freight flows. Eur. Transp. Res. Rev. **13**, 33 (2021)

Milakis, D., Athanasopoulos, K.: What about people in cycle network planning? applying participative multi-criteria gis analysis in the case of the athens metropolitan cycle network. J. Transp. Geogr. **35**, 120–129 (2014)

Mitra, R., Nash, S.: Can the built environment explain gender gap in cycling? an exploration of university students' travel behavior in toronto, canada. Int. J. Sustain. Transp. **13**(2), 138–147 (2019)

Molloy, J., Schmid, B., Becker, F., and Axhausen, K. W. (2019). mixl: An open-source R package for estimating complex choice models on large datasets. *Arbeitsberichte Verkehrs-und Raumplanung*, 1408. Publisher: IVT, ETH Zurich

Nehme, E.K., Pérez, A., Ranjit, N., Amick, B.C., Kohl, H.W.: Sociodemographic factors, population density, and bicycling for transportation in the united states. J. Phys. Act. Health **13**(1), 36–43 (2016)

Pearson, L., Dipnall, J., Gabbe, B., Braaf, S., White, S., Backhouse, M., Beck, B.: The potential for bike riding across entire cities: quantifying spatial variation in interest in bike riding. J. Transp. Health **24**, 101290 (2022)

Pearson, L., Gabbe, B., Reeder, S., Beck, B.: Barriers and enablers of bike riding for transport and recreational purposes in australia. J. Transp. Health **28**, 101538 (2023)

Pucher, J., Dill, J., and Handy, S. (2010). Infrastructure, programs, and policies to increase bicycling: An international review. *Preventive Medicine*, 50:S106–S125. 01323

Reggiani, G., Van Oijen, T., Hamedmoghadam, H., Daamen, W., Vu, H.L., Hoogendoorn, S.: Understanding bikeability: a methodology to assess urban networks. Transportation **49**(3), 897–925 (2022)

Rupi, F., Poliziani, C., Schweizer, J.: Data-driven bicycle network analysis based on traditional counting methods and gps traces from smartphone. ISPRS Int. J. Geo Inf. **8**(8), 322 (2019)

Sallis, J.F., Conway, T.L., Dillon, L.I., Frank, L.D., Adams, M.A., Cain, K.L., Saelens, B.E.: Environmental and demographic correlates of bicycling. Prev. Med. **57**(5), 456–460 (2013)

Scott, D.M., Lu, W., Brown, M.J.: Route choice of bike share users: Leveraging gps data to derive choice sets. J. Transp. Geogr. **90**, 102903 (2021)

Shaw, C., Russell, M., Keall, M., MacBride-Stewart, S., Wild, K., Reeves, D., Bentley, R., Woodward, A.: Beyond the bicycle: Seeing the context of the gender gap in cycling. J. Transp. Health **18**, 100871 (2020)

Steele, W.: Who governs australia's metropolitan regions? Australian Planner **56**(2), 59–64 (2020)

Sun, Z., Lorscheid, I., Millington, J.D., Lauf, S., Magliocca, N.R., Groeneveld, J., Balbi, S., Nolzen, H., Müller, B., Schulze, J., et al.: Simple or complicated agent-based models? a complicated issue. Environ. Model. Software **86**, 56–67 (2016)

Thompson, J., Wijnands, J.S., Savino, G., Lawrence, B., Stevenson, M.: Estimating the safety benefit of separated cycling infrastructure adjusted for behavioral adaptation among drivers; an application of agent-based modelling. Transport. Res. F: Traffic Psychol. Behav. **49**, 18–28 (2017)

Trapp, G.S., Giles-Corti, B., Christian, H.E., Bulsara, M., Timperio, A.F., McCormack, G.R., Villaneuva, K.P.: On your bike! a cross-sectional study of the individual, social and environmental correlates of cycling to school. Int. J. Behav. Nutr. Phys. Act. **8**, 1–10 (2011)

Venter, Z.S., Gundersen, V., Scott, S.L., Barton, D.N.: Bias and precision of crowdsourced recreational activity data from strava. Landsc. Urban Plan. **232**, 104686 (2023)

Victoria State Government: Plan Melbourne 2017–2050. Technical report, Department of Environment, Land, Water and Planning, Melbourne (2017)

Wałdykowski, P., Adamczyk, J., Dorotkiewicz, M.: Sustainable urban transport-why a fast investment in a complete cycling network is most profitable for a city. Sustainability **14**(1), 119 (2021)

Willis, D.P., Manaugh, K., El-Geneidy, A.: Cycling under influence: Summarizing the influence of perceptions, attitudes, habits, and social environments on cycling for transportation. Int. J. Sustain. Transp. **9**(8), 565–579 (2015)

Zahran, S., Brody, S.D., Maghelal, P., Prelog, A., Lacy, M.: Cycling and walking: Explaining the spatial distribution of healthy modes of transportation in the united states. Transp. Res. Part D: Transp. Environ. **13**(7), 462–470 (2008)

Zhao, P.: The impact of the built environment on bicycle commuting: Evidence from beijing. Urban studies **51**(5), 1019–1037 (2014)

Ziemke, D., Metzler, S., Nagel, K.: Bicycle traffic and its interaction with motorized traffic in an agent-based transport simulation framework. Futur. Gener. Comput. Syst. **97**, 30–40 (2019)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Afshin Jafari[1] · **Steve Pemberton**[1] · **Dhirendra Singh**[2,3] · **Tayebeh Saghapour**[1] · **Alan Both**[1] · **Lucy Gunn**[1] · **Billie Giles-Corti**[1,4,5]

✉ Afshin Jafari
afshin.jafari@rmit.edu.au

1 Centre for Urban Research, RMIT University, Melbourne, Australia

2 School of Computing Technologies, RMIT University, Melbourne, Australia

3 CSIRO's Data61, Clayton, Australia

4 The Kids Research Institute Australia, Perth, Australia

5 The University of Western Australia, Perth, Australia