ORIGINAL RESEARCH

Abstracting mobility flows from bike-sharing systems

Fabio Kon[1,2] · Éderson Cássio Ferreira[1] · Higor Amario de Souza[1] ·
Fábio Duarte[2,3] · Paolo Santi[2,4] · Carlo Ratti[2]

Abstract

Bicycling has grown significantly in the past ten years. In some regions, the implementation of large-scale bike-sharing systems and improved cycling infrastructure are two of the factors enabling this growth. An increase in non-motorized modes of transportation makes our cities more human, decreases pollution, traffic, and improves quality of life. In many cities around the world, urban planners and policymakers are looking at cycling as a sustainable way of improving urban mobility. Although bike-sharing systems generate abundant data about their users' travel habits, most cities still rely on traditional tools and methods for planning and policy-making. Recent technological advances enable the collection and analysis of large amounts of data about urban mobility, which can serve as a solid basis for evidence-based policy-making. In this paper, we introduce a novel analytical method that can be used to process millions of bike-sharing trips and analyze bike-sharing mobility, abstracting relevant mobility flows across specific urban areas. Backed by a visualization platform, this method provides a comprehensive set of analytical tools to support public authorities in making data-driven policy and planning decisions. This paper illustrates the use of the method with a case study of the Greater Boston bike-sharing system and, as a result, presents new findings about that particular system. Finally, an assessment with expert users showed that this method and tool were considered very useful, relatively easy to use and that they intend to adopt the tool in the near future.

Keywords Bike-sharing · Mobility · Data science · Visualization · Open source software · Application case study

---

✉ Fabio Kon
  kon@ime.usp.br

[1] Department of Computer Science, University of São Paulo, São Paulo, Brazil

[2] Senseable City Lab, Massachusetts Institute of Technology, Cambridge, USA

[3] Pontifícia Universidade Católica do Paraná, Cambridge, USA

[4] Istituto di Informatica e Telematica del CNR, Pisa, Italy

1 Introduction

The use of bicycles for short trips (from 1 to 5 km) in medium to large cities for commuting, occasional, and leisure trips presents multiple proven benefits at the global, local, and personal level. In *global* terms, manufacturing a car creates much more CO_2 emissions and consumes much more energy than manufacturing a bicycle, and the same can be said about bike versus car trips (Shaheen and Lipman 2007; Weiss et al. 2015). Concerning *local* benefits to the city, an increase in cycling trips in substitution of motorized trips helps mitigating traffic congestion, decreasing air and noise pollution, and the amount of required parking space (Sælensminde 2004). In addition, it also brings several *personal* benefits to both mental and physical health (Oja et al. 2011). Research shows that commuting to work on a bike also presents advantages concerning other active modes of transportation, such as walking, since its higher cardio-respiratory intensity is associated with health benefits (Shephard 2008; Hoevenaar-Blom et al. 2011). However, both pedestrians and cyclists are more exposed to injuries than car or transit passengers (Margie Peden et al. 2004). In the case of cycling, the risk is aggravated when dedicated bike lanes are not available (Lusk et al. 2011).

Bike-sharing systems (BSS) were first deployed at a small scale in Amsterdam in 1965 and grew slowly throughout the following decades (Demaio 2009; Jan Ploeger 2020). The first generation of BBS is characterized by unlocked bicycles spread over in public places for free use (Shaheen et al. 2010). In 1991, a second generation, offering a few hundred coin-operated bikes, was born in small cities in Denmark. In 1995, the first large-scale second generation BSS was deployed in Copenhagen, known as Bycyklen or City Bikes. In 1996, a third generation, based on magnetic cards and several technological advances, was initiated in England and continued to evolve within the following years. But it was only when Lyon, in 2005, and later Paris, in 2007, made their wide deployments of several thousand shared bikes that these systems started to become known worldwide (Demaio 2009; Jan Ploeger 2020). In a few years, similar programs spread throughout all the continents with a massive growth in the number and scale of systems.

At the turn of the century, a fourth generation of BSS started to appear, requiring no docking stations; they are called free-floating or dockless BSS. A precise characterization of the fourth generation is not consensual, but most authors tend to consider dockless BSS as the fourth generation (Fishman 2020). The early implementations of those recent dockless BSS were very small scale and based on phone calls or SMS. With the growth of the mobile internet and the popularization of smartphones, fourth-generation BSS systems with tens of thousands of bicycles started to be deployed in 2015.

Shaheen et al. (2010) argued that fourth-generation BSS are mainly distinguished by advances in related technologies and by expanding the ways to use BSS. Among these features are the offer of electric bicycles, mobile and solar-powered dock stations, and integration with other transportation modes. Thus, they do not consider dockless systems as a mandatory feature to define the fourth generation of BSS.

An exponential growth of bike-sharing has been observed in both developed and developing countries, in large and small, dense, and sprawling cities (Yanocha et al. 2018). Arguably, no other mobility system has grown so fast: from less than 100 bicycles in 2002, the bike-sharing fleet had grown to more than 600,000 in 2013. Recent estimates mention that there are more than 18 million (Richter 2018) or more than 20 million (Fishman 2020) shared bikes worldwide. A precise estimate is not possible since most dockless systems are run by private companies that do not share their numbers.

In late 2019, a market research report predicted that the bike-sharing services market was expected to double its size, from USD 2.7 billion in 2018 to USD 5 billion by 2025 (PSI 2019). With the COVID-19 crisis, it is now challenging to predict what will happen in the next few years, but several cities worldwide are incentivizing cycling as a way to promote social distancing, trying to lower the number of commuters in crowded buses and subways (Huet 2020).

This growth of bike-sharing over the last decade might be explained by the fact that bike-sharing systems encapsulate global values that are slowly becoming a consensus among policymakers, such as the promotion of healthier lifestyles and the acknowledgment of the negative environmental impacts of motorized transportation. BSS are also made possible by global technological tendencies, such as the widespread use of information technologies to amass geolocated travel data that boosts fleet management and transportation planning (Duarte 2016). Besides, bike-sharing systems also present a few advantages over cycling with one's own bike, such as (1) not requiring private parking space both at home and at the destination, (2) not requiring to carry a bike inside a bus, subway, or train when connecting to another transportation mode, (3) not having to worry about maintenance or theft, and (4) being able to use the bike in just one trip during the day or in a few disconnected trips, i.e., not requiring a round-trip.

One of the main arguments for implementing BSS is that they provide an effective alternative for the first- and last-mile problem, mainly when integrated with public transport (Yanocha et al. 2018; Hong et al. 2016). According to Yanocha et al. (2018), 28 million bike-sharing trips were performed in the USA in 2016. However, this still represents a very small proportion of the trips made in major cities. Data from the USA Department of Transportation's 2017 National Household Travel Survey indicate that 35% of car trips in the USA were shorter than 2 miles, and almost half of them were less than 3 miles–a distance that could usually be covered by cycling (NHTS 2018). Thus, there are plenty of opportunities for the expansion of such systems both to new cities and within cities that already have limited BSS implementations. BSS have been implemented around the world in ad hoc manners, with little scientific, evidence-based planning. The complex dynamics of such systems and their interaction with city life and other means of transportation are not yet fully understood. There are multiple business models, and public and private forms of funding BSS. Within the past few years, several BSS companies have gone bankrupt, and most cities worldwide are still reluctant to consider bike-sharing as an integral part of their mobility portfolio. Thus, it is not yet clear what would be a successful business model to support BSS, and a better understanding of its dynamics is desirable.

In this paper, we propose a method and introduce an associated open source tool, BikeScience, for the analysis of bike-sharing mobility flows as a means to support the monitoring, understanding, and planning of the cycling infrastructure of cities. Our approach differs from previous works into the field of BSS data analysis as the proposed methodology involves dividing the city into regular grids of different sizes to aggregate thousands of bike trips and derive flows at different levels of detail. As input, our method uses the origin–destination (OD) trip data from bike-sharing systems (or other sources such as OD surveys). It produces a series of metrics and visualizations that help city officials, transport authorities, urban planners, and BSS companies to better understand bike mobility patterns and derive actions to improve cycling policies and infrastructure.

The tool implementing the proposed method is distributed as a collection of open source Python libraries and Jupyter notebooks available at https://gitlab.com/interscity/bike-science. A Jupyter notebook is an interactive software apparatus based on a specific programming language (Python in our case) and related data science libraries. It contains detailed associated documentation and can be customized by programmers to different contexts. Based on the same set of Python libraries, our group also developed an interactive web-version of the tool, which does not require the local installation of the notebooks.

After an evaluation of the tool by Transport officials in the city of São Paulo, they decided to adopt the tool for internal use at the city Transportation Authority. The experts found that the abstraction of bike flows is useful in planning the ongoing growth of the city cycling infrastructure. The tool provides them with evidence for implementing the city policy to make cycling an alternative way of transportation. Before using BikeScience, they did not have any robust mechanism to incorporate quantitative data into their planning toolset. They used to plan the cycling infrastructure based on their personal knowledge and also by performing public consultations with citizens.

We have worked along with the transportation planners to develop new analyses in BikeScience. For example, we are implementing a bike flow analysis using data from the São Paulo's 2017 origin–destination survey (OD17). This survey contains trips made by citizens with their own bicycles in a broader (metropolitan) area around the city, which is complementary to BSS data. We are also working together to cross-analyze bike flows and bicycle counters. The aim is to combine those data to update the OD17 bicycle trip information, since this travel survey is performed only every ten years.

In the remainder of this paper, we discuss the main factors influencing urban cycling (Sect. 2.1), related work (Sect. 2.2), and then present requirements (Sect. 3) and our method (Sect. 4). We describe the Greater Boston case study (Sect. 5) to illustrate the use of our method. Finally, we show how we evaluated our contribution (Sect. 6) and present our conclusions (Sect. 7).

2 Literature review

We now discuss existing literature that is relevant to our work. We start with the factors that influence urban cycling in general and bike-sharing in particular. Then, we discuss related works on bike-sharing data analysis and how this work differs from them.

2.1 Influential factors

Previous work by van Acker et al. (2010) concluded that travel behavior derives from locational behavior and activity behavior but is also influenced by lifestyle, perceptions, attitudes, and preferences. A cross-sectional study of six cities in the US showed strong effects of individual attitudes and physical and social environment factors on bicycle ownership and use (Handy et al. 2010).

As depicted in Fig. 1, a wide variety of cultural, social, geographical, economic, and infrastructural factors influence the cycling activity in contemporary cities (Fishman 2016). In addition to other means of transportation, two relevant factors are the available supply of bikes (via bike-sharing systems and other sources) and the existing city infrastructure to support cycling. The major stakeholders influencing the growth of cycling in the past two decades have been city governments, investors, bike-sharing companies, and, of course, users (cyclists) (Fishman 2016; Zhang et al. 2015; Raux et al. 2017). The latter use bikes for commuting, travel on employers' business, leisure, tourism, and everyday errands (Ricci 2015; Fishman 2016).

Four common types of infrastructure that provide comfort and safety for cyclists are (in decreasing order): (1) protected bike lanes (aka physically-separated bike lanes or segregated cycle tracks), (2) bike-pedestrian shared trails, (3) painted bike lanes, and (4) painted lanes shared with cars (aka sharrows) (Cambridge 2014). The quality they provide to users is proportional to their cost, which can vary
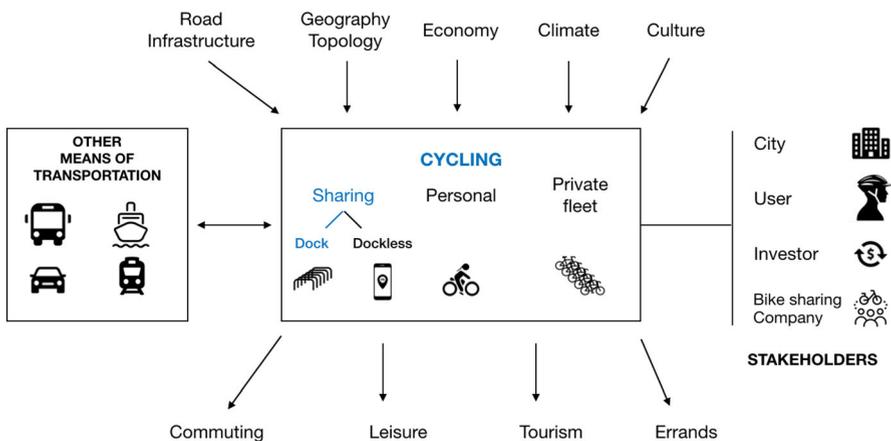

Fig. 1 Factors influencing urban cycling

substantially (Bushell et al. 2013). The average cost per mile of a cycling track/lane in the USA is around $130,000.

Thus, understanding where the major flows of cyclists are located within a city is the first step in providing urban planners with the knowledge required to draw a good mobility plan for urban cycling. However, this alone is not enough as the availability/lack of a proper cycling infrastructure can directly influence the existence/absence of bike flows. It is also relevant to understand which factors explain the existence of these flows and also which flows were expected based on those factors but are not observed in reality. For example, the proximity of the subway and a large density of jobs and businesses are natural sources of bike trips. Although Wang et al. (2016) did not study bike flows, they analyzed the usage of bike stations, showing how it is influenced by sociodemographic, built environment, transportation infrastructure, and economic variables. Maldonado-Hinarejos et al. (2014) studied the role of individual attitudes and perceptions in predicting the demand for cycling. Frade and Ribeiro (2014) proposed a framework to predict bike-sharing demand based on urban space topography and travel demand.

Beyond the general benefits of cycling, BSS are a means of facilitating access to cycling when carrying a private bicycle is not so practical. When BSS is seen as part of the transportation infrastructure of a city, it helps to fill the gaps left by the conventional public transportation systems. Successful BSS implementations are responsible for a large portion of bicycle trips in large cities where they are present [e.g., 40% of the bicycle trips in Paris (Fishman 2020)]. For a part of the population, BSS can provide advantages in terms of convenience, time-saving, and cost in relation to other means of transportation (Chen et al. 2020a).

On the other hand, BSS may also bring disadvantages (Fishman 2020; Gu et al. 2019). For example, the fact that BSS and associated cycling infrastructures reduce space for car parking is considered a disadvantage for a significant portion of car users. However, providing more BSS stations and cycling infrastructure can discourage car trips, a strategy that has been adopted in several cities (e.g., Boston, Amsterdam, Melbourne) and is often seen as an advantage by contemporary urban planners. Another disadvantage mentioned in the literature is the urban clutter that is sometimes produced by bikes left on the curbs and sidewalks, mainly in the case of large dockless deployments (Gu et al. 2019). But, compared to the cluttered streets and curbsides caused by cars, this problem is really small and can be managed.

2.2 Related work

Most previous work on BSS data analysis focuses on investigating usage patterns of individual dock stations (O'Brien et al. 2014; Faghih-Imani et al. 2014; NHTS 2018; Sarkar et al. 2015; Wang et al. 2016; Faghih-Imani et al. 2017; Duran-Rodas et al. 2019) and/or on analyzing the overall usage of BSS in multiple cities concerning urban features (Zhao et al. 2014; Duran-Rodas et al. 2019). Such studies do not investigate the movements from one place to another, i.e., the origin–destination pairs of bike trips.

Research that addressed mobility patterns often focuses on the flows from one individual station to the other (Corcoran et al. 2014; Beecham and Wood 2014;

Zhao et al. 2015; Zhang et al. 2018), which can provide interesting insights on the punctual dynamics of the system. The problem is that stations usually are distributed unevenly across the city, and each station is a too fine granularity that does not provide an overall picture of city mobility dynamics for the urban planner. Wood et al. (2011) proposed an elegant method for visualizing flows that resemble our visualization technique; but their work visualizes flows from and to each station, without spatial aggregation.

More recently, there have been studies on dockless systems. Chen et al. (2020b) conducted an extensive survey with users to analyze the factors that promote the use of dockless BSS. Shen et al. (2018) use GPS data from dockless trips to analyze the distribution of trips across the regions of the city and its relationship with characteristics of the surrounding environment and weather; they do not consider the trajectories of the trips nor the mobility flows that are generated.

Médard de Chardon et al. (2017) collected data from 75 BSS in different countries to analyze the number of trips per bike per day as a comparison of performance and success. Their main conclusion is that increasing system size does not increase performance. Duran-Rodas et al. (2019) analyzed data from different cities to identify factors that lead to the popularity of docking stations. They found that population, distance to city center, leisure-related establishments, and transport infrastructure influence ridership for particular stations. They did not look into flows.

There are a few initial works that looked at mobility flows using bike-sharing data. Zhou (2015) used a clustering algorithm to group together flows connecting stations in Chicago, identifying 378 relevant flows in the city for the year of 2014; this is an interesting approach but showing so many unstructured flows to the user does not support policy-making adequately; in addition, the computational complexity of the clustering algorithm might hinder the method interactivity. Similar to our work, Beecham et al. (2014) propose a visual analytic approach for analyzing bike-sharing data. However, their focus is on automatically labeling commuting journeys, based on a spatial analysis of travel behaviors. Their method is then used to study commuter behavior.

Jie Bao, Tianfu He, et al. propose a data-driven approach to develop bike lane construction plans. Based on available budget, construction convenience, and bike lane utilization constraints, they propose greedy-based heuristics for deriving cycling infrastructure construction plans Bao et al. (2017); He et al. (2020). They analyze one month of trajectory data from Mobike users, a dockless system, to identify the road segments that are used the most. Then, they use a greedy network expansion algorithm to generate suggestions for where to build bike lanes. This is a complementary approach that could be used in combination with the one proposed in this paper.

Different from all previous work, we propose here a novel method, and associated open source toolkit, to identify and analyze the most relevant bike flows within a city. Different from the results published in the literature, our method presents the information in a structured way, aggregating the trips from multiple stations by regions, and providing rankings at multiple levels of granularity that can be selected interactively by the urban planner or BSS operator. Different from

previous approaches to flow analysis, our tool is capable of analyzing over a million trips in a few seconds on a commodity laptop computer, providing a high degree of interactivity for the user.

3 Requirements

To define the requirements for an effective methodology and tool for analyzing bike-sharing data, providing a useful instrument for policy-making, we follow a two-step process. First, we wanted to provide features that would be useful for transport planners from city governments and bike-sharing companies; for that, we interviewed professionals in the cities of Boston (USA), São Paulo (Brazil), and Beijing (China) from June to December 2018. Second, we conducted an extensive literature review of scientific papers published in international journals in the field of bike-sharing data analysis. This literature review showed that there was little previous research on identifying and analyzing cycling flows. Expert users, mainly Transport Engineers from city governments, confirmed that this was really a gap in the existing tooling for them. Thus, we decided to focus our research on abstracting bike-sharing flows and their analysis, which was a gap both in scientific research and in practical tools. Expert users also noted that it would be useful to have a means to quickly extract descriptive statistics data for bike usage, selecting specific periods, time of day, age and gender, and to plot this geographically on a map, contrasting the data with existing transport and cycling infrastructure. This last point has been performed in previous scientific work, but no integrated and flexible tool was easily available to practitioners.

In particular, transportation authorities were especially interested in visualizations that would contrast the most important mobility flows (extracted from bike-sharing and OD survey data) with the existing and planned city cycling infrastructure (mainly segregated cycle tracks and painted bike lanes).

Thus, the major requirements for our method and tool were:

1. Easy importing and processing (with the press of a button) of raw data with millions of trips from bike-sharing or OD survey databases.
2. Selecting specific periods of the dataset, specific time of day, day type (weekday/weekend), gender, age.
3. Analyzing trip distance, duration, speed.
4. Identifying the most important bike-sharing mobility hubs in the city.
5. Finding the most significant mobility flows at different granularity levels including at least, (1) metropolitan area, (2) group of neighborhoods and (3) local area inside a neighborhood.
6. Visualizing, in an interactive map, the most relevant flows, superimposing them with the underlying transportation infrastructure (cycling infrastructure, metro stations, and roads).
7. Contrasting the recommended bike routes for the most relevant flows with existing cycling infrastructure.

8. All these features should be available in a user-friendly, interactive tool that should provide answers in a few seconds (ideally, no more than 10 or 20 s for each analysis).

We now define the methodology we created to fulfill these requirements and describe a concrete case study.

4 Methodology

Our approach is based on dividing the city into homogeneous regions by using a uniform grid and computing the number of bike trips from one grid cell to another. We then draw directed arcs to show the flow direction and adjust the origin and end point of flows according to a weighted average based on the station usage for that specific flow. By manipulating the various input parameters of this method, one can analyze different characteristics of specific mobility patterns. To simplify computation and the discussion, we use rectangular grid cells in this paper, but the method can also be used with hexagon grids (Birch et al. 2007).

We now describe in detail the input datasets our method requires and, then, the method components.

4.1 Datasets

Currently, hundreds of cities around the world provide information about their bike-sharing systems as open data available on public web sites. Although most of them contain similar information, they often use slightly different formats. The General Bikeshare Feed Specification[1] is a widely-used open standard for real-time bike availability on stations, but it does not cover the actual trips, which is the data we use in our analysis. Thus, when starting to work with data from a new city or from a new system within a city, an initial step of adapting the format is required. Typically, trip data come in multiple CSV files, each one containing 1 month of data and follows a format similar to the one depicted in Table 1, which lays out the most recent format for the Boston Bluebikes system.

For the use case described in Sect. 5, we processed information about 8,447,044 trips from the Boston Bluebikes system available at https://www.bluebikes.com/system-data, covering the period from 2011 to 2018. The data was stored in different formats, with a few distinct patterns in different years; so an initial step of data cleaning and standardization to the format in Table 1 was required. For some of the analysis described in this paper, to have a better picture of the current state of the system, we focus on the most recent 2-year period, from November 2016 to October 2018, which contains over 3 million trips.

For each trip, we use the location and time of origin and destination as well as the gender, age, and type of user (subscriber or casual customer). The gender and age are only available for subscribers of the system (annual or monthly member,

---

[1] https://github.com/NABSA/gbfs.

Table 1 Sample data format specifying one trip

field	trip duration	start time	start station id + name + position	end station id + name + position
format	seconds	timestamp	int + string + LatLong	int + string + LatLong

field	bike id	usertype	birth year	gender
format	int	Subscriber/Customer	int	(F, M, Unknown)

accounting for 78% of the trips), not for casual customers (single trip or day pass users, accounting for 22% of the trips).

4.2 The method

The method we propose is composed of a basic procedure (Flow Abstraction) and three additional steps (Input Filtering, Metrics Selection, and Results Presentation), which enable the execution of a large diversity of spatial and numerical analysis. We detail the procedures in the following.

4.2.1 Fundamental procedure: flow abstraction

The goal here is to extract *abstract flows* from collections of trips. First, we take the BSS coverage area and partition it into a homogeneous two-dimensional grid containing $N \times M$ rectangular cells (Fig. 4). Then, for every possible (*origin*, *destination*) pair of cells, we compute how many trips were performed starting at *origin* and ending at *destination*, and these counts will compose the flows. However, a naïve implementation of this procedure could lead to bad performance, which would make the computation of thousands of flows over millions of trips to take several minutes. This would hinder the interactivity of a software based on our method, harming its effectiveness as an interactive tool to support public policy and analysis.

Bike trips are characterized by the geographical coordinates of their station start and end points. Thus, we need to match those points to the grid cells where they belong to by performing some geographical processing. By working with large datasets in a tabular format, we took advantage of optimized modern data science libraries such as *Pandas* for numerical processing and *GeoPandas* for geographical processing.[2]

The procedure and its operations are described in Algorithms 1 and 2. Given a collection of trips $\mathcal{T}$, a geographic grid $\mathcal{G}_{M,N}$ of size M x N cells over the area under study, and a collection of bike stations $\mathcal{S}$, the algorithm computes a collection of

---

[2] By representing origin and destination cells (i,j) as columns in a data frame structure, it is fast to sum up the trips using the aggregation facilities in the Pandas library. To further speed up the process, we maintain a separated file with the station coordinates and performed GeoPandas spatial joins between stations and grid cells, not between trips and cells.

flows $\mathcal{F}$. Each trip $t \in \mathcal{T}$ is a tuple containing the information in Table 1. Each cell $c_{i,j} \in \mathcal{G}$ is a rectangle. Each bike station $s \in \mathcal{S}$ is a geographical point. All geographical data are described in terms of latitudes and longitudes. In the output produced by the algorithm, each flow $f \in \mathcal{F}$ is a tuple that contains the origin and destination cells of that flow and its weight. All (c_s, c_e) cell pairs in $\mathcal{G} \times \mathcal{G}$ become a flow if there is at least one trip t that starts in c_s and ends in c_e. The weight of a flow is defined as the number of trips abstracted by that flow.

In Algorithm 1, *Spatial_Join*[3] performs the geospatial preprocessing step to relate the cells in $\mathcal{G}$ to the bike stations in $\mathcal{S}$ (line 2). The result, *stations_and_cells*, is a set of pairs (*station*, *cell*) s.t. *station* is located within the *cell* area.

Line 3 computes the number of trips that start in an origin station and end in a destination station by using a *GroupBy.Count* call. *GroupBy* extracts all pairs of origin and destination stations and *Count* reduces the collection to one tuple per group, computing the number of trips in each group and assigning the count to a new attribute, *trip_count*.

In lines 4 and 5, we obtain the corresponding grid origin and destination cells for each pair of stations extracted in line 3. The *Join* operation matches two collections $\mathcal{A}$ and $\mathcal{B}$ and returns all $(a, b) \in \mathcal{A} \times \mathcal{B}$ with a specified attribute in common. The result of line 3 is joined twice against the result of line 2, for both origin and destination stations, matching the ids of the stations. The end result, stored in *od_stations*, is a set of tuples, each containing the pair of stations, the number of trips between them, and the pair of cells in which the stations are located.

In line 6, we use Algorithm 2 to calculate the center of mass for the origin and destination cells of each flow. The center of mass is a geographical point computed as the average of the station locations, weighted by the number of trips to or from each station for one particular flow. Each flow has two centers of mass, one in the origin cell and another in the destination. Thus, the centers of mass tend to be closer to stations with more trips in that flow. In this manner, we abstract an origin and destination for each flow that is more representative than adopting, for example, the center of the grid cell.

Line 7 computes the number of trips with start and end stations that belong to the same origin and destination cells. Line 8 joins the centers of mass with each pair of origin and destination cells. At this point, $\mathcal{F}$ is the collection of flows, including their origin, destination, and weight, i.e., the output of Algorithm 2.

In Algorithm 2, line 2 normalizes the number of trips for all pairs of origin and destination stations between 0 and 1 and stores that value in the *trip_count* column. Lines 3 to 6 compute the latitude and longitude for each pair of origin and destination stations weighted by the normalized *trip_count* value. Line 7 groups all pairs of origin and destination cells (flows) and sums their weighted coordinates to obtain the average coordinate points that will represent the origin and destination for each flow.

---

[3] The function *Spatial_Join* is available in geospatial tools like GeoPandas (http://geopandas.org). The functions *GroupBy*, *Join*, *Normalize*, *Sum*, and *Count* are available in Data Science tools such as Pandas (https://pandas.pydata.org).

The time complexity of Algorithm 1 is $O(|\mathcal{T}|)$. The *Spatial_Join* operation (line 2) is $O(|\mathcal{G}| + |\mathcal{S}|)$, which is linear since each station belongs to only one cell. *Join* operations in Pandas can be made on indexes, passing over the dataframe rows at once, resulting in $O(|\mathcal{G}|)$ or ($O(|\mathcal{S}|)$), depending on their sizes. *GroupBy.Count* (line 3) is a sequence of *split-apply-combine* operations over $\mathcal{T}$, which are done in a single pass over the data, resulting in $O(|\mathcal{T}|)$. Lines 4 and 5 have a similar complexity behavior as line 1, since they are *Join* operations between *stations_and_grid_cells* ($\mathcal{X}$) and $\mathcal{S}$, which would be $O(|\mathcal{X}|)$ or $O(|\mathcal{S}|)$. Line 6 depends on Algorithm 2, which starts with a *GroupBy.Normalize* operation (line 2) over the rows ($\mathcal{R}$) of the OD dataset and ends with a *GroupBy.Sum* operation (line 7) over the rows of the *centers* dataset. The other lines of Algorithm 2 are $O(1)$. As the sizes of the *centers* and OD datasets are equal, Algorithm 2 is $O(|\mathcal{R}|)$. Back to Algorithm 1, the *GroupBy.Sum* operation in line 7 is also $O(|\mathcal{R}|)$. Line 8 is $O(|\mathcal{F}|)$ because the set of mass centers is $O(|\mathcal{G}|)$ while $\mathcal{F}$ is $O(|\mathcal{G}|^2)$. All other steps in Algorithm 1 are linear operations. Thus, as $\mathcal{T} \gg \mathcal{G} \wedge \mathcal{S} \wedge \mathcal{X} \wedge \mathcal{R} \wedge \mathcal{F}$, the time complexity of Algorithm 1 is linear with the size of the bike trips dataset.

Algorithm 1 Flow abstraction

1: **procedure** ABSTRACTFLOWS($\mathcal{T}, \mathcal{G}, \mathcal{S}$)
2:     $stations\_and\_cells \leftarrow$ SPATIAL_JOIN($\mathcal{G}, \mathcal{S}$)
3:     $od\_station\_counts \leftarrow \mathcal{T}$.GROUPBY($start\_station\_id, end\_station\_id$)
          .COUNT($new\_column = trip\_count$)
4:     $od\_stations \leftarrow od\_station\_counts$.JOIN($stations\_and\_cells, start\_station\_id$)
5:     $od\_stations \leftarrow od\_stations$.JOIN($stations\_and\_cells, end\_station\_id$)
6:     $mass\_centers \leftarrow$ CENTEROFMASS($od\_stations$)
7:     $\mathcal{F} \leftarrow od\_stations$.GROUPBY($start\_cell, end\_cell$).SUM($trip\_count$)
8:     $\mathcal{F} \leftarrow F$.JOIN($mass\_centers, start\_cell, end\_cell$)
9:     **return** $\mathcal{F}$
10: **end procedure**

Algorithm 2 Center of Mass

1: **procedure** CENTEROFMASS($od\_stations$)
2:     $centers \quad \leftarrow \quad od\_stations$.GROUPBY($start\_cell, end\_cell$).NORMALIZE($column \quad = \quad trip\_count$)
3:     $centers.start\_station\_lat \leftarrow centers.start\_station\_lat * centers.trip\_count$
4:     $centers.start\_station\_lon \leftarrow centers.start\_station\_lon * centers.trip\_count$
5:     $centers.end\_station\_lat \leftarrow centers.end\_station\_lat * centers.trip\_count$
6:     $centers.end\_station\_lon \leftarrow centers.end\_station\_lon * centers.trip\_count$
7:     $centers \leftarrow centers$.GROUPBY($start\_cell, end\_cell$).SUM($start\_station\_lat,$
          $start\_station\_lon, end\_station\_lat, end\_station\_lon$)
8:     **return** $centers$
9: **end procedure**

Algorithm 3 Separation of flows into tiers

1: **procedure** SEPARATEFLOWSINTOTIERS($\mathcal{F}, max\_tiers$)
2:     $total\_trips \leftarrow \text{SUM}(\mathcal{F}.trip\_count)$
3:     $top\_flow \leftarrow \text{MAX}(\mathcal{F}.trip\_count)$
4:     $total\_flows \leftarrow \text{COUNT}(\mathcal{F})$
5:     $tiers \leftarrow$ empty list
6:     $flows\_and\_tiers \leftarrow$ empty list
7:     $tier\_num \leftarrow max\_tiers$
8:     $partial \leftarrow 0$
9:     $num\_flows \leftarrow 0$
10:    $previous\_value \leftarrow 0$
11:    $flows \leftarrow \mathcal{F}.\text{SORTDESCENDINGBY}(column = trip\_count)$
12:    **for** $f \in flows$ **do**
13:       $flows\_and\_tiers \leftarrow flows\_and\_tiers \cup \{(f, tier\_num)\}$
14:       $num\_flows \leftarrow num\_flows + 1$
15:       $partial \leftarrow partial + f.trip\_count$
16:       **if** $partial >= (max\_tiers - (tier\_num - 1)) \times total\_trips/max\_tiers$ **then**
17:          $flows\_perc \leftarrow num\_flows/total\_flows$
18:          $tiers \leftarrow tiers \cup \{(tier\_num, top\_flow, previous\_value, num\_flows, flows\_perc)\}$
19:          $tier\_num \leftarrow tier\_num - 1$
20:          $num\_flows \leftarrow 0$
21:          $top\_flow \leftarrow f.trip\_count$
22:       **end if**
23:       $previous\_value \leftarrow f.trip\_count$
24:    **end for**
25:    **return** $tiers, flows\_and\_tiers$
26: **end procedure**

4.2.2 Step 1: Input filtering

Depending on the specific analysis under consideration, we filter the data with respect to different parameters such as age, gender, trip duration, distance, average speed, date, and time. Workdays present similar patterns among themselves but they differ significantly from weekends and holidays, so we normally treat these classes separately. Within a single day, we investigate three different time periods: morning peak (from 7:00 to 10:00), lunchtime (from 11:00 to 14:00), and afternoon peak (from 17:00 to 20:00) as their patterns differ significantly.

Thus, to produce the input $\mathcal{T}$ for Algorithm 1, one needs to define a filter $\mathcal{FIL} = \langle p, dt, tod, gs, gen, age, dis, dur, spe, ti \rangle$ specifying a collection of up to 10 features that will be used to select the trips that will compose $\mathcal{T}$. The ten features are described in Table 2.

The raw amount of flows within a city is very large. Showing all of them to a user is overwhelming and does not allow any reasonable analysis. Thus, using Algorithm 3, we divide $\mathcal{F}$, the collection of flows, into four tiers, each one containing 25% of the trips. As we will show in Sect. 5.4, the top tiers tend to be composed of few flows that contain a large number of trips, and the bottom tiers contain a large number of flows with very few trips. The time complexity of Algorithm 3 is $O(|\mathcal{F}|)$, thus it is linear with the number of flows ($\mathcal{F}$) among the grid cells. Statements in lines 2, 3, and 4 are aggregation operations over $\mathcal{F}$. Lines 12-24 compose a loop sequence that iterates over $\mathcal{F}$. The other statements of the algorithm are $O(1)$.

Table 2 Set of features used to filter a collection of trips

Feature symbol	Feature name	Typical value range	Default value
p	Period	1 month, 1 year	All
dt	Day type	All, workday, weekend/holiday	All
tod	Time of day	All, morning, lunch, afternoon	All
gs	Grid size	From 10×10 to 100×100	20×20
gen	Gender	All, male, female	All
age	Age	From 20 to 70	All
dis	Distance	From 500 m to 10 km	All
dur	Duration	From 3 min to 60 min	All
spe	Speed	From 10 km/h to 25 km/h	All
ti	Tier	All, 4, 3, 2, 1	All

4.2.3 Step 2: metrics selection

After applying the filtering procedure in Step 1, one would typically have a set of hundreds of thousands or a few million trips. In Step 2, we need to define which metrics will be used in the analysis. The metrics can be related directly to the trips (e.g., number of trips, percentage of trips to a total, distance, median, speed, percentage of short trips vs. long trips, distribution to a variable, etc.). The same metrics can be calculated over trips in different tiers, different time periods, according to different values of categorical variables, such as gender, to analyze specific effects.

Alternatively, the metrics can be related to flows. In this case, the Flow Abstraction procedure is executed in the output of the filtering procedure to generate the set of flows $\mathcal{F}$. Then, one computes the metrics on the flows (e.g., number of flows, number of trips per flow, distribution of flows across tiers, etc.).

For the analysis involving distances and speed, we estimate the road distance between two stations using the GraphHopper API (http://graphhopper.com) over OpenStreetMap. In particular, we use the bike mode route planner, which provides bike-friendly routes. GraphHopper uses the A* and Dijkstra algorithms for pathfinding and relies on OpenStreetMap data for optimizing paths for bikes. Based on data from OpenStreetMap, the algorithm takes into consideration the surface, type of highway, and elevation. It also puts a penalty on roads without a cycle path and an even higher penalty for roads with dangerous places like tunnels or when cars can go faster than 50 km/h. Finally, the algorithm gives preference to roads that are part of an official cycle network or where the road is a pure bicycle-street. These suggested bike routes are typically 30% longer than the Euclidean distance, on average. We used the API to build a matrix containing n^2 cells, where n ($n = 308$ in the case study described in Sect. 5) is the number of dock stations, each cell containing the distance between two stations. To improve data collection time, we batch routing requests in groups of 80, the maximum amount allowed by the API, i.e., each request returns an 80×80 matrix with the 360 distances. We then cache the

collected information locally in a CSV file so that we only need to call the GraphHopper API again when the set of stations change (e.g., when analyzing a new city or when stations are added or moved).

4.2.4 Step 3: presentation of results

After computing the metrics, the results can be presented either in a numeric format (e.g., in tables or graphs) or in a spatial form, typically in maps. Numeric data can be displayed in a wide variety of formats promoting human cognitive understanding, for example, histograms (Fig. 2), line graphs (Fig. 3), and violin plots (Fig. 8).

Spatial presentation of results typically plots different values on a geographical map of the area under study. In our case, we are particularly interested in plotting the flows, while relating them to different aspects of the city, including cycling infrastructure, bus stops, subway stops, bike stations, suggested routes (from tools such as Google Maps or GraphHopper), or sociodemographic variables from the Census.

A crucial element in visualization is the capacity of clearly depicting the flows against a background map with the least visual confusion as possible, avoiding clashes among multiple flows. To that end, we plot each flow using curved arrows from origin to destination in which the width of the arrow is proportional to the number of trips in that flow, with respect to the other flows drawn in the same diagram. The use of curved arrows instead of straight lines minimizes the superposition of flows connecting different regions. It is also useful to visualize flows connecting the same two grid cells but in opposite directions; for that, we compute the arrow curvature to be upward in flows moving west and downward for flows moving east.
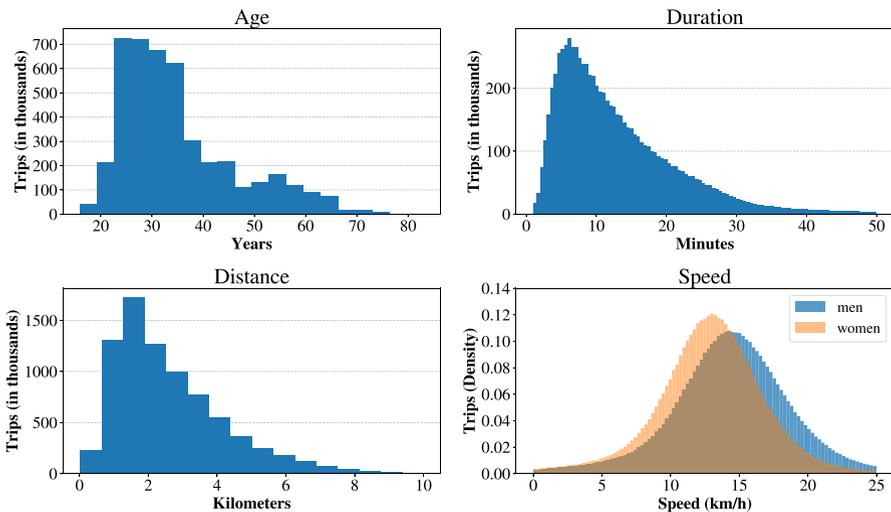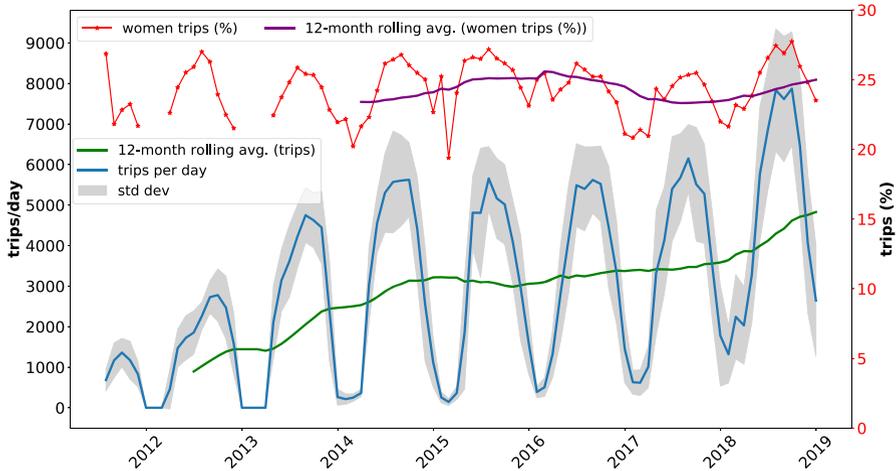

Fig. 2 Descriptive statistics for Boston Bluebikes data

Fig. 3 Evolution of trips from 2011 to 2019

The exact location of the start and end points of the arrow is computed separately for each flow as the average of the coordinates of the dock stations in that grid cell, weighted according to the number of trips for each station for that particular flow (as described in Algorithm 2).

A concrete implementation of this method was developed in Jupyter notebooks leveraging python libraries for data analysis and geovisualization. To ensure the reproducibility of our results, all the source code and datasets used in this paper are available at https://gitlab.com/interscity/bike-science. We welcome the contribution of other researchers and programmers willing to help extending the scarce body of open source tools for bike data analysis; we will provide the necessary support for contributors.

5 The Boston case study

To illustrate the method presented in this paper, we use 8 years of data from the Boston BlueBikes bike-sharing system as a case study. Boston (Boston Transportation Department 2017) is a relatively bike-friendly city, having received a silver medal award (LAB 2018) from the League of American Bicyclists in 2017. From 2007 to 2014, the bicycle lane mileage in Boston went from 0.03 to 92 miles, with a decrease in bicycle accidents around 14% per year (Pedroso et al. 2016). Boston's original bike-sharing system, Hubway, was launched in 2011, and it has been growing since then. In 2018, its name changed to BlueBikes and it now has over 1800 bicycles and 308 fixed stations across Boston, Brookline, Cambridge, and Somerville; each station has from 10 to 47 docks. With the BikeScience open source tool introduced herein, we analyzed 8.45 million bike trips since the inception of the bike-sharing program. Fig. 2 helps us understanding usage patterns extracted from BlueBikes data for 2011 to 2019. We used BikeScience to produce the age, and trip

distance, duration, and speed histograms depicted in the figure. The tool presents these descriptive statistics in a graphical form, either in its notebooks or in an interactive web application.

Trip duration follows a log-normal distribution with a median of 10 min and with 75% of the trips taking under 16 min. The speed follows a Student's t-distribution, with men riding slightly faster than women.

Figure 3 shows the evolution of the total number of trips per day for the entire system. One can see both the strong seasonal effects of the typical harsh winter in Boston and the overall tendency for an increase in usage over the 8 years (confirmed by the 12-month rolling average drawn in green in the figure).

5.1 The gender gap

The percentage of female subscribers using the system (drawn in red in Fig. 3) shows not only that men use bike-sharing more frequently but that the difference increases during winter when roads become less safe with snow and ice. The women/men ratio was lowest in the Winter of 2015, with only 19.4% female users in February. The highest relative presence of women so far was in September 2018, when 27.7% of subscriber trips were made by women. The cities of Boston, Cambridge, and Somerville have improved the quality and extension of their cycling infrastructure in recent years (see Table 3 with the most recent data from the Boston Department of Transportation). As women feel more comfortable and secure in the cycling tracks, the gender gap tends to decrease (Lusk et al. 2011; Garrard et al. 2012).

To verify whether we could identify a trend in the gender gap in bike usage, we computed the 12-month rolling average percentage of female users, shown in the purple line. In the past 5 years, there is no clear trend: the 12-month rolling average percentage of female users has been oscillating between 22 and 26%. There is a steady increase in the ratio of women in the past 18 months, but it is still too soon to know whether this will be a long-term trend in the Boston area.

Thus, we can see that, if we desire to have fairer access to cycling concerning gender, specific policies targeting women should be implemented. These might involve both improvements in the urban infrastructure and educational campaigns for cyclists and motor vehicle drivers.

Table 3 Miles of cycling infrastructure per year (Boston)	Infrastructure type	2014	2015	2016
	Protected bike lanes	1.2	2.2	2.2
	Buffered bike lanes	2.9	3.5	4.6
	Bike lanes	48.0	51.3	51.9
	Sharrows	29.5	33.4	35.1
	Total	81.6	89.4	93.8

5.2 Grid sizes and levels of detail

When applying the flow abstraction procedure described in Algorithm 1, one needs
to define the desired level of detail to perform the analysis. By using different grid
sizes, an urban planner can understand the most relevant flows within the city at
different levels of granularity. As depicted in Fig. 4, the Boston BlueBikes stations
are unevenly spaced across the area, thus analyzing the flows from station to station
does not yield a balanced analysis of urban flows. The flow abstraction algorithm
solves that by reasoning in terms of geographical regions (grid cells).

Dividing the BlueBikes coverage area into large regions using a 10x10 grid leads
to cells that are a little larger than a square mile (1.8 × 1.8 km). This enables us to
visualize the major flows across the cities that compose the area as well as flows
across distant neighborhoods within a city. Figure 5b presents the morning peak
flows with this resolution, showing a significant west-to-east movement in
Cambridge and flows from Brookline, Somerville, and Boston towards Cambridge.
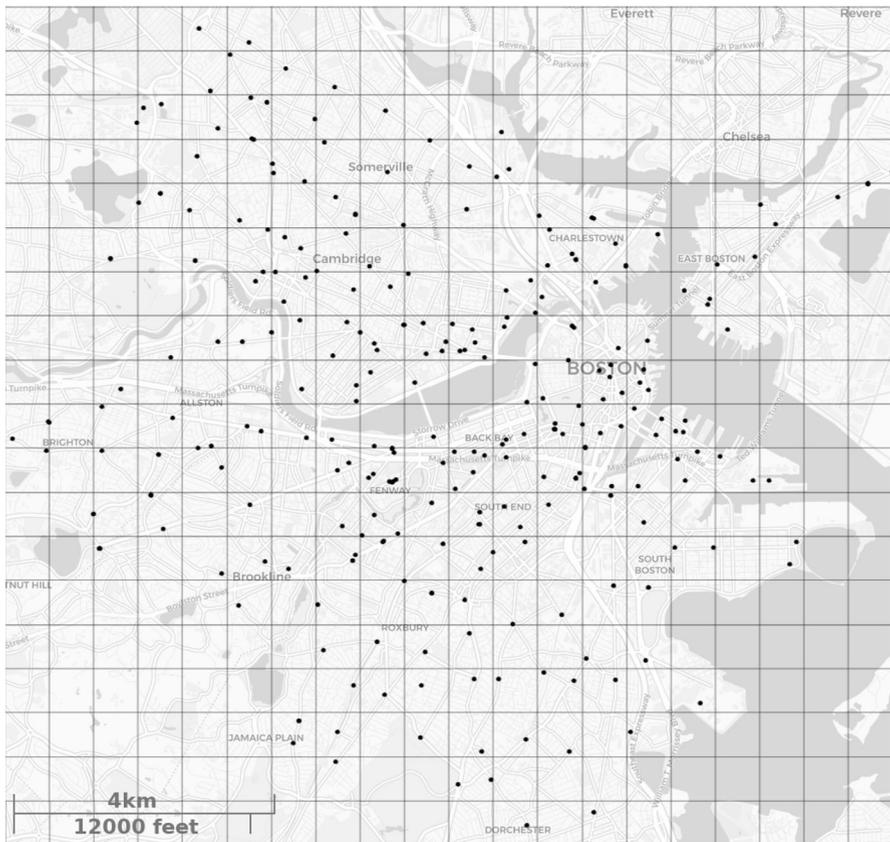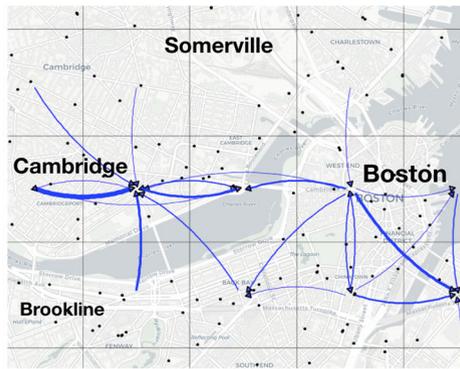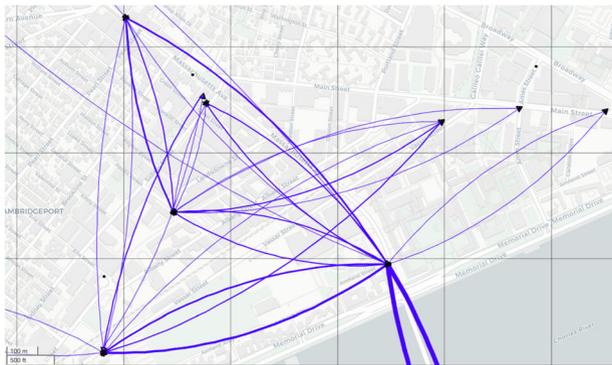

Fig. 4 Bluebikes coverage area and 20 × 20 grid partitioning
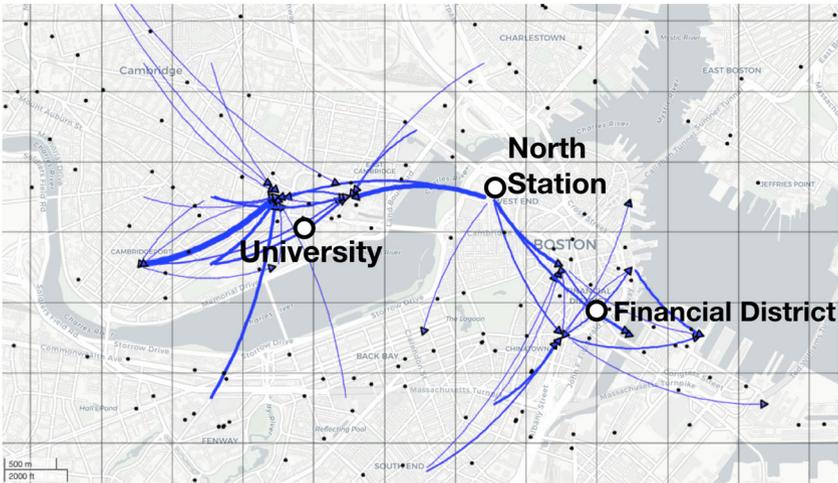
(a) 20x20 grid (morning rush hour)

(b) 10x10 grid - flows across cities

(c) 40x40 grid - local flows within a neighborhood in Cambridge

Fig. 5 Bike flows at different levels of granularity

A more detailed view is provided by a grid of size 20 × 20, which shows the flows from a region the size of a typical neighborhood to the others; in Fig. 5a, one can visualize, for example, strong flows connecting Boston's North Train Station to a major university and to the Financial District in the morning rush hour. Finally, a 40 × 40 grid, with square cells of 450 m, is capable of showing the local bicycle flows within neighborhoods (Fig. 5c); with this granularity, most of the flows represent single connections between two specific stations and a public manager or BSS operator can identify bottlenecks or underutilized stations. In BikeScience, the user can click on specific flows to get additional information, such as the number of trips that flow represents. In this analysis, we discard trips that start and end within the same grid cell, which represents 3.5% of the flows in a 40 × 40 grid, increasing to 5.5% and 16.1% for the 20 × 20 grid and 10 × 10 grid, respectively.

Thus, city officials and BSS operators can use the method to visualize flows with various grid sizes to analyze different aspects of mobility flows within the city, from a macro-level across cities in a metropolitan area to a micro-level within specific neighborhoods of a city.

Using different scales to aggregate areas may lead to different visualization patterns (Wong 2009). Such a difference is a consequence of the Modifiable Areal Unit Problem (MAUP) (Openshaw and Taylor 1979), in which the results of an analysis may vary due to the type of spatial aggregation used to model a geographical space. In BikeScience, users can change the grid size to analyze areas of different sizes. Large grid cells (e.g., a 10 × 10 grid) will concentrate more bike stations and more trips per grid cell. Also, the trips that occur inside a given cell are discarded from the abstraction of flows. Conversely, smaller grid cells lead to fewer stations and fewer bike trips per grid cell.

To mitigate MAUP-type issues, our method relies on plotting the arrows representing the flows with the origin and destination not on the center of the grid cell, but rather in the geographical center of mass of the trips associated with that flow. This helps to alleviate abrupt changes in the flows related to small changes in the grid location and size. According to our experience with BikeScience, the overall changes in the flow patterns when the grid size and position changes are not very drastic. Nevertheless, an analyst using the tool must be aware of the MAUP problem and, when in doubt, play with different grid sizes. Alternatively,

| (a) 10x10 grid | (b) 20x20 grid | (c) 30x30 grid |
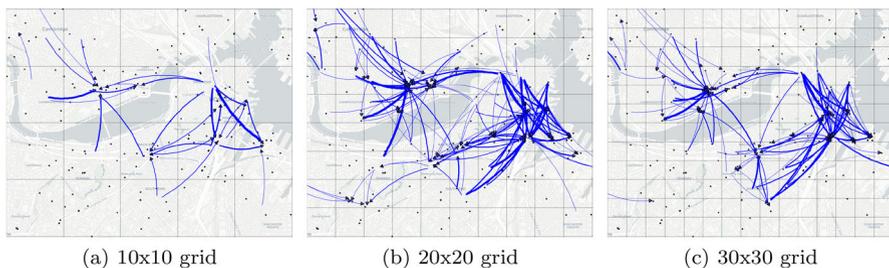
Fig. 6 Variation of bike flows at different grid levels

BikeScience also offers the option of hexagon-based grids, which might present advantages in certain situations (Birch et al. 2007).

Figure 6 shows the variation of flows using rectangular 10×10, 20×20, and 30×30 grids. The flows in the figure represent morning workday trips from July 2011 to July 2019. By decreasing the grid cell size, it is possible to see an increase in the number of flows, which brings more detailed information about how trips occur in smaller areas.

5.3 Time periods

Mobility flows vary greatly across seasons (as shown in Fig. 3) and times of the day. Workday flows also present patterns that are different from the ones on weekends and holidays. As we described, our method, and its associated BikeScience tool, lets users define filters to specify which kind of data they want to analyze. In particular, three filters are relevant here: period (p), day type (dt), i.e., workday, weekend, or holiday, and time of day (tod).

A city office promoting tourism might want to understand the mobility patterns across touristic areas on weekends (Fig. 5c), while a transportation authority might be interested in workday commuting patterns during the morning (Fig. 9a) and afternoon (Fig. 9b). Finally, a business alliance in a certain commercial region might get insights from analyzing lunchtime flows in its area (Fig. 9c).

5.4 Tiers

As discussed in Sect. 4.2, simply displaying all flows to a user, as performed in previous related work, is overwhelming and does not allow a reasonable analysis. For example, with a 10×10 grid for morning workday trips, from November 2016 to October 2018, the Boston area has almost 2 thousand different BSS flows. To show this information in a comprehensible manner, our method divides the flows in tiers. Table 4 and Fig. 7 illustrate this approach by dividing the flows in 4 tiers, each one containing 25% of the trips.

The top quarter (tier 4) contains only 1% of the total flows but represent 25% of the trips while the bottom quarter (tier 1) contains 1730 different flows. From Table 4 and the violin plots in Fig. 8, one can see that the most frequently used flows in the top tier aggregate up to 14,303 trips and tend to relate to shorter trips. As Fig. 8 shows, the least used flows (in tier 1) tend to have longer trips.

The top quarter typically has only a few tens of flows (18 in our example) and indicates areas in which good cycling infrastructure such as segregated, protected bike lanes would benefit a large number of trips. The second top quarter typically has several tens of flows (45 in our example), and indicates areas where the city could, for example, develop infrastructures such as buffered lanes or painted bike lanes on shared streets. Of course, this analysis would need to consider other factors such as the currently available infrastructure and the proximity of other means of public transportation, and potential demand generated by residential, business, and educational centers.

Table 4 Distribution of trips across four tiers

Tier	Trips per flow	# flows	% flows	% trips	Median distance (km)
4	5111 to 14303	18	1	25	2.15
3	1797 to 4851	45	2	25	2.41
2	618 to 1787	118	6	25	3.24
1	Up to 617	1730	91	25	4.41

Tier	Median duration (m:s)	% short trips < 1.5 km	% long trips > 4 km
4	10:05	23	2
3	11:31	20	9
2	14:37	12	29
1	23:33	7	55

According to the data in Table 4, the city can take care of half of the bike trips by investing in only 3% of the flows. Alternatively, developing infrastructure for 9% of the flows would benefit 75% of the trips. The bottom tier 1 is spread over 1730 different flows; thus, most cities would not have financial means to invest in specific cycling infrastructure for that, but it is an area in which overall education of the population concerning cycling could bring safety to cyclists, pedestrians, and drivers.

Table 5 shows the number of flows and the percentage of flows in each tier as one varies the grid size. There are more flows in all tiers as we make grid cells smaller. However, the percentage of trip flows in each tier does not vary very much. Tier 4 contains nearly 1% of all flows for all grid sizes, confirming the results shown in Table 4.

5.5 Trip lengths

Applying filters related to the distances covered by trips enables useful examination of different classes of trips concerning their length. Urban planners investigating the use of bikes for longer trips may inspect, for instance, trips longer than 4 km. The Boston analysis shows that most long trips occur in the morning or afternoon peak hours (in opposite directions), with few such trips at lunchtime. This indicates a high usage of such trips by commuters who are willing to start or end their day with a long trip (requiring 15–30 min). However, as shown in Fig. 9a, b, while the morning trips are concentrated on fewer flows (likely to be from home to work), the afternoon trips are spread across a wide variety of flows, evidencing that bikers probably undertake different kinds of activities after the work hours (gym, grocery stores, happy hour with friends, dinner, etc.).
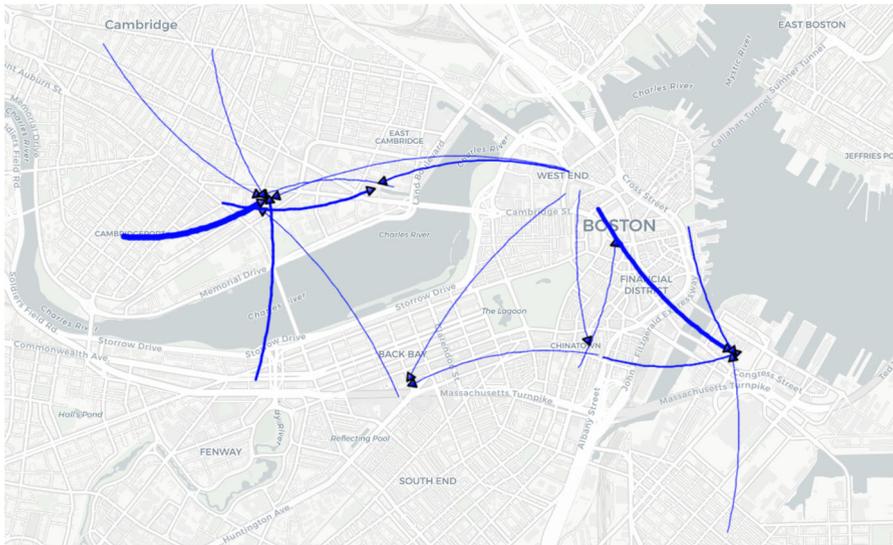
(a) Tier 4 (add protected bike lanes?)
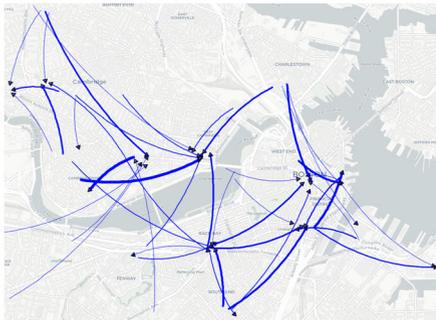
(b) Tier 3 (add painted bike lanes?)   (c) Tier 2 (paint shared car-bike lanes?)

Fig. 7 Workday morning for the top 3 tiers—each tier represents 25% of the trips

Regarding short trips, the patterns are different. There is a significant amount of them at lunchtime and Fig. 9c shows a few clear clusters throughout the city.

5.6 Mobility hubs

In addition to the mobility flows, another relevant information is what parts of the city are the major hubs initiating or ending bike trips. Our method can be used to analyze this issue by computing the number of trips per region instead of abstracting the flows, as illustrated in Fig. 10. This map shows dark green markers in the top regions of the city where bike trips start, and dark red markers in the top regions where trips end. Light green and light red markers point to second-tier hubs. This information can be used by city officials as one of the elements to guide where
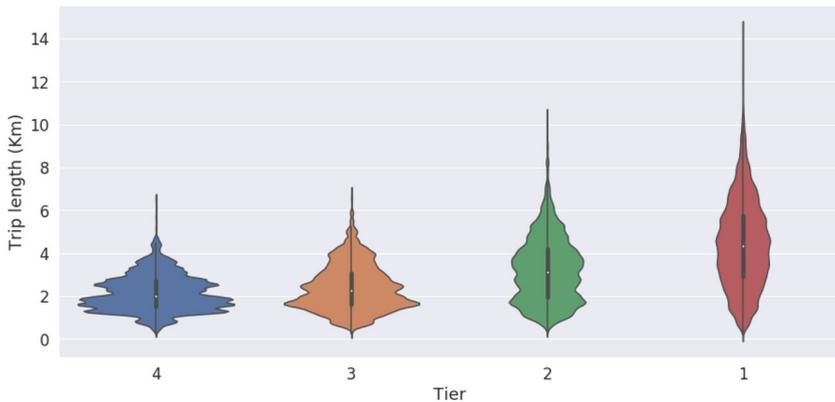
Fig. 8 Trip length distribution across tiers

**Table 5** Distribution of flows across four tiers for different grid sizes

Tier	10 × 10		20 × 20		30 × 30		40 × 40	
	# flows	% flows	# flows	% flows	# flows	% flows	# flows	% flows
4	18	1	89	1	132	1	212	1
3	45	2	252	3	420	3	667	4
2	118	6	710	8	1152	8	1749	10
1	1730	91	8347	89	11,976	88	15,481	85

investments should be made in local infrastructure for bicycles. The information might also be valuable for businesses offering bike-related services and products. While bike-sharing users do not need to maintain their bicycles, the presence of a large number of trips by bike-sharing subscribers in a specific area might be a good proxy for the overall presence of cyclists, i.e., including privately-owned bikes.

An alternative way to visualize this information, showing a more global view of the entire system is via conventional heat maps such as those depicted in Fig. 11, which shows the overall distribution of morning trip starts and ends in October 2018 workdays.

5.7 Speeders

The proposed method can be easily used to perform analysis focused on specific situations such as detecting evidence of rider reckless behavior. The most common reason for cycling accidents and fatalities is to be hit by a car (Coleman and Mizenko 2018). Although the car driver is at fault many times, according to the US Department of Transportation, from 2010 to 2015, the most common bicyclist action before fatal accidents was the cyclist failure to yield right-of-way (34.9%)

(a) Long trips - Morning

(b) Long trips - Afternoon

(c) Short trips - Lunch time
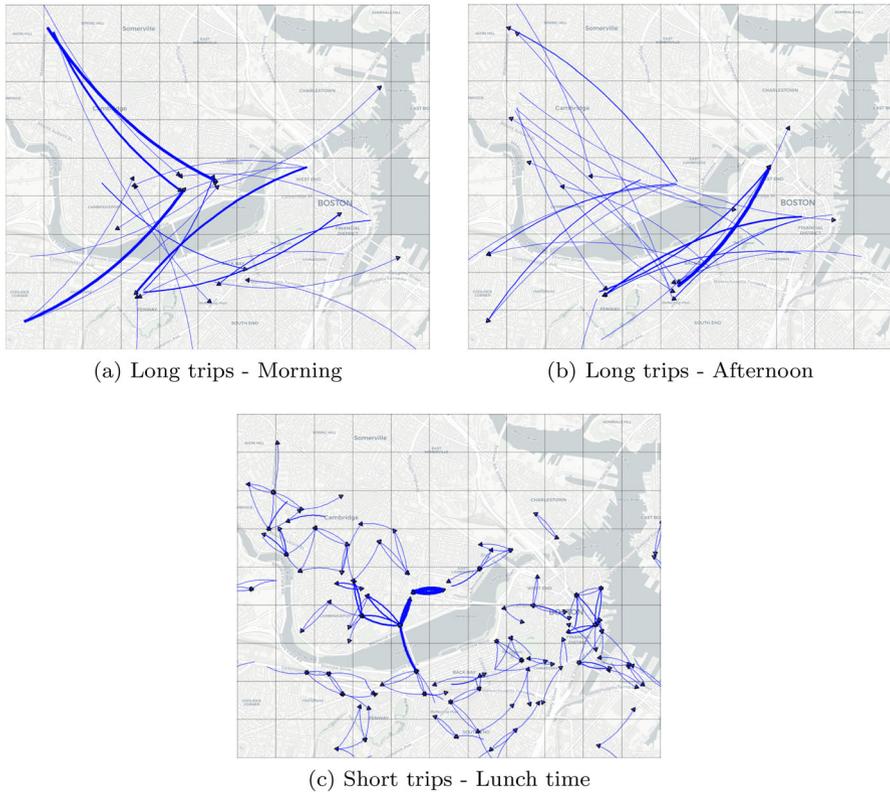
Fig. 9 Long trips (> 4 km) and short trips (< 1 km)

(Coleman and Mizenko 2018). A city government, then, may wish to develop an educational campaign to decrease the number of cyclists that ride their bikes dangerously fast.

Using our method, one can easily produce an analysis of speeders. Of course, this would be limited to BSS users and not to the general cyclist population; but it might provide useful insights for the design of an educational campaign. For example, in the Fast-Cyclists notebook available with the tool, we selected the trips whose average speed was over 20 km/h. Given that the average speed for all trips is 13 km/h and that only the 4.2% of the trips are above 20 km/h, we considered these trips to be likely associated with cyclists riding dangerously fast. We can then draw a profile of these speeders:

- 89% are men while only 11% are women;
- 50% of them are between 21 and 32 years old. They are present in all age ranges under 52, but the age in which people have a higher tendency to drive dangerously fast is 25–30;
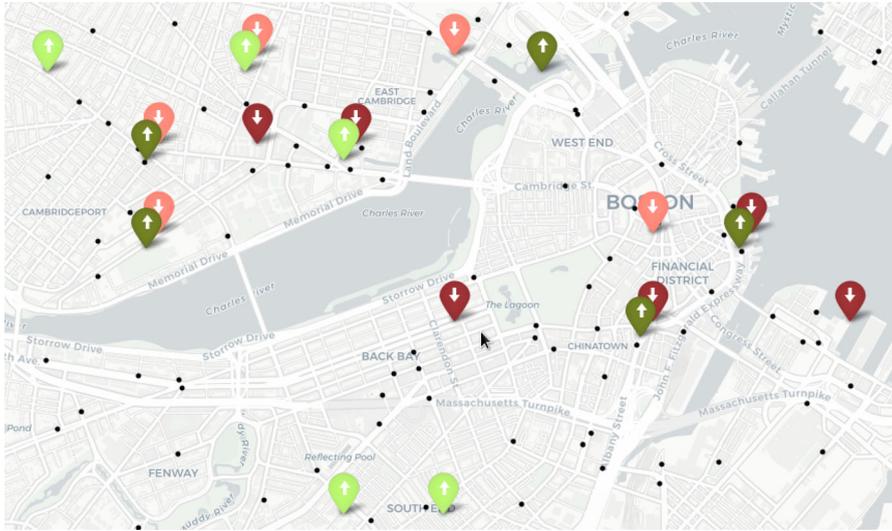
Fig. 10 Mobility Hubs: top trip start hubs (green), top trip end hubs (red), second-tier start/end hubs (light green/light red)

- Speedy trips length is 20% longer than average (they might speed because they need to go farther away). Speedy trips duration is half of the average of all trips (they want to get there quickly);
- A subscriber (usually, a resident of the area) is 4.6 times more likely to be a speeder than an infrequent customer (usually, a tourist).

Such a profile can help the city to develop an educational campaign targeted at a specific group. Besides, the geographical analysis of the flows shows on the map where most speeding trips are often located, which can help with both the educational campaign and law enforcement.

5.8 Additional data sources

An issue that must be carefully considered with our method is that the existence/absence of a flow might be induced by the existence/absence of the cycling infrastructure. Thus, urban planners must analyze the flows identified by the method in conjunction with other data sources. Currently, BikeScience includes the cycling infrastructure, bus stops, subway and train stations, as well as suggested bike routes. Transport authorities and urban planners can also visualize the flows with respect to different geographical and administrative regions of the city, which might be more intuitive for them; for example, Fig. 12 depicts flows over the official city neighborhoods.

By cross-analyzing the most relevant flows with respect to existing cycling infrastructure, one can identify areas with significant flows that lack appropriate cycling infrastructure or vice-versa. Our method lets urban planners compare the
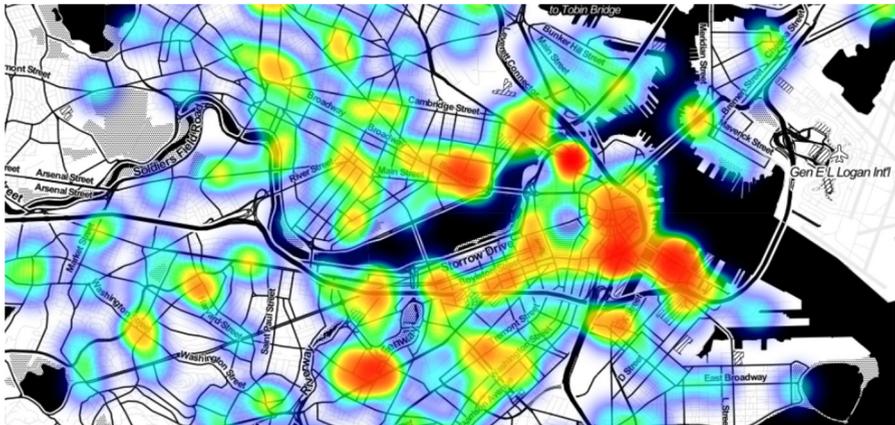
(a) Morning trip start heatmap (red zones around train and subway stations)
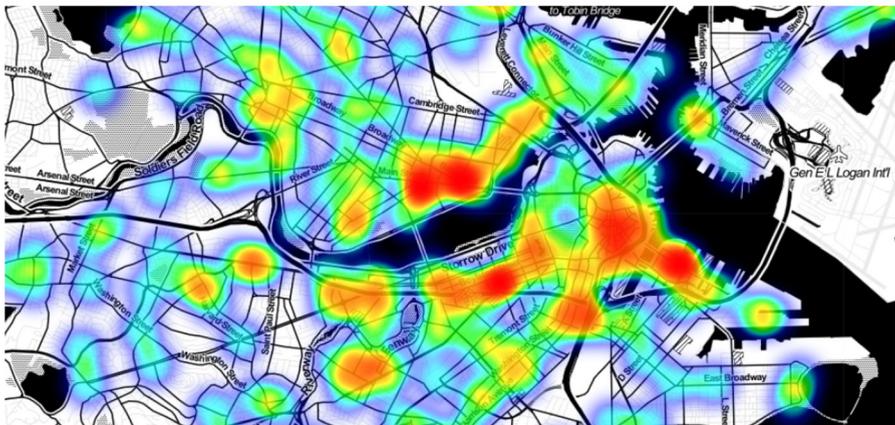
(b) Morning trip end heatmap (red zones around universities and business centers)

Fig. 11 Oct. 2018 trip start/end concentration

most relevant flows with existing infrastructure. For example, Fig. 13a depicts Boston downtown morning flows against protected bike lanes (in red), bike trails shared with pedestrians in parks (in green), and car-shared bike lanes, sharrows, (in orange). An urban planner can easily identify flows that are well served by cycling infrastructure and flows that are not. By selecting a specific flow, BikeScience draws a new map showing the recommended bike route for that flow (according to the GraphHopper online API), so that the urban planner can analyze in detail how well the infrastructure serves that route. Figure 13b depicts (in a dashed black line) the recommended bike route, which, in this case, does not offer good infrastructure for
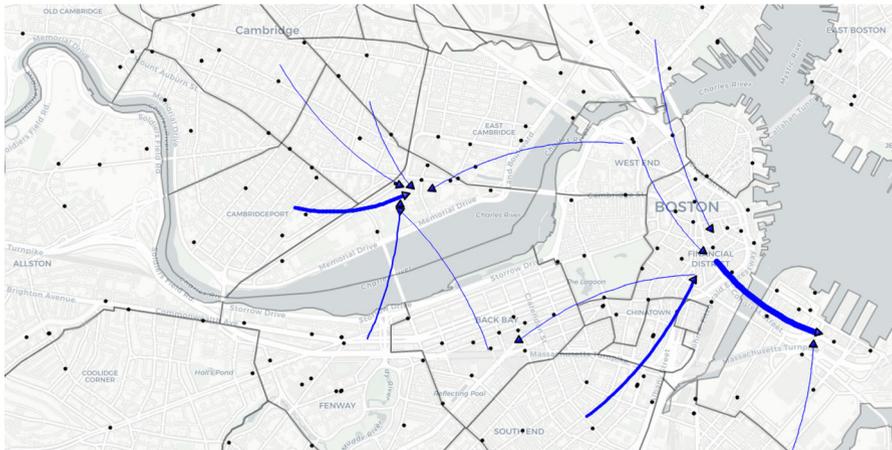
**Fig. 12** Major morning flows across Great Boston official neighborhoods

bikers, who are forced to share the street with cars, with no protection for most of the route.

We are currently working to include further geolocated demographic and socioeconomic information about the population, jobs, schools, universities, museums, parks, and entertainment centers into the BikeScience tool. Adding new data variables to the system is relatively easy and requires little effort from a python programmer. On the other hand, presenting relevant information to the end-user in a useful way requires careful consideration. This work will provide urban planners with the possibility of correlating mobility flows with the city infrastructure and socioeconomic variables. For example, if an area with many jobs is not far from a subway station but does not present a significant flow of bikes, this might highlight an important deficiency of cycling infrastructure in that region.

5.9 Environmental influential factors

To understand the influential environmental factors of bike-sharing trips in Boston, we studied the relationship between trips and the points of interest (POIs). We gathered POI data using the Google Places API (https://developers.google.com/places/web-service/intro). There are more than 136 place types available in the API, from which we identified 97 that occur in the Boston metro area. Among the gathered data, we found place types for food (e.g., cafes, restaurants, bars), arts (e.g., museums, libraries), services (e.g., accounting, post offices, ATMs), transportation (e.g., train stations, subway stations), shopping (e.g., clothing stores, department stores), education (e.g., schools, universities), health (e.g., doctors, dentists, clinics, hospitals), and sports (e.g., gym). We collected geolocated information about 48,347 POIs in the area covered by Boston Bluebikes.

We applied the *Pearson's product correlation coefficient* (r) (Cohen 1977) to assess the relationship between bike-sharing trips and POIs. The *r* values vary from
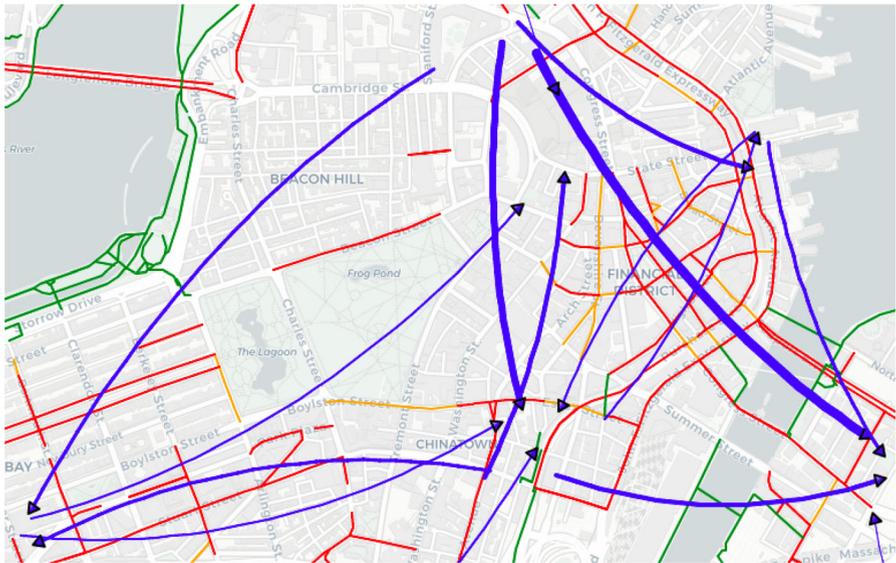
(a) Morning downtown flows
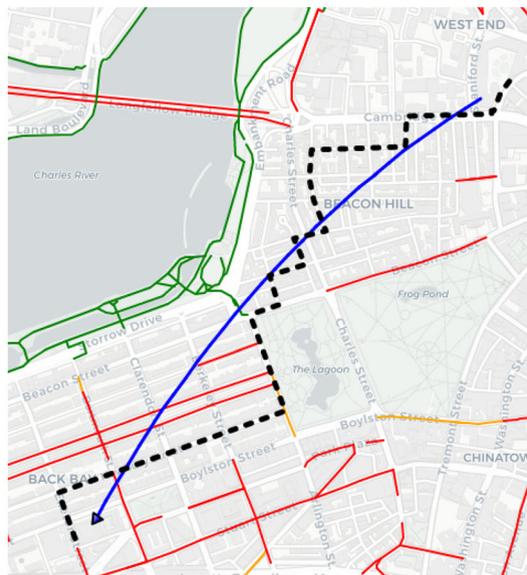
(b) Recommended route (dashed black line)

Fig. 13 Protected bike lanes (red), bike-pedestrian trails (green), and sharrows (orange) against bike-sharing flows (blue)

− 1 to 1, where *-1* means a total negative correlation, *0* means no correlation, and *1* means a total positive correlation.

To proceed with the correlation analysis, we computed each POI type per grid cell and correlated it to the number of trips that start (or end) in the correspondent cell. When a user picks up or returns a bicycle in a station, s/he does not necessarily intend to reach a POI within that grid cell. S/he may intend to go to a place that is located in a neighboring cell. To consider this fact, we used *Geodesic buffering* (Flater 2011) to distribute the weight of the POIs to their nearest cells. This buffer is a circle that represents an influence area from the center of a cell. The circle radius is less than the distance between the center of two neighboring cells. If a POI is in the influence area of only one cell, it belongs entirely to that cell. However, if a POI is in the reach area of two cells, both cells will have half of that POI. The same rationale applies to POIs that area shared by three or four cells (Flater 2011).

We found that several POI types are highly correlated with bike-sharing trips. In most scenarios we evaluated, more than 50 POIs have correlation values above 0.5. Table 6 shows the 10 POIs with the highest correlations and 10 POIs with the lowest correlation values. By looking at this table, one can have a clear view on what kind of places attract a large number of bike trips.

The results are consistent through distinct scenarios, i.e., the POIs with high correlation values were roughly the same for different grid sizes and periods of a day. Cafes, parking, and restaurants are the most correlated to bike-sharing trips in almost all cases. We observe an increase in the correlation of the destination of trips to areas with more restaurants in the late afternoon/early evening.

These results indicate that machine learning algorithms could be able to predict the occurrence of flows based on this kind of data. In fact, a preliminary study we conducted (https://gitlab.com/interscity/bike-science/-/tree/master/ml-models) showed that a random forest model based on POIs, demographic data extracted from the US Census, and weather data was capable of predicting correctly most of the Boston flows with high accuracy. However, further research must be conducted for this model to be generalizable to other cities.

Table 6 POIs with best and worst correlations

High-correlation POIs	Correlation	Low-correlation POIs	Correlation
Cafe	0.83	Bowling alley	0.01
Parking	0.80	Cemetery	0.02
Restaurant	0.79	Car wash	0.02
Bar	0.78	Mosque	− 0.03
Museum	0.78	Painter	0.04
Gym	0.77	Synagogue	− 0.04
Post office	0.76	Hindu temple	0.07
Real estate agency	0.76	Stadium	0.07
Lodging	0.76	Pet store	0.07
Finance	0.76	Storage	0.09

6 Evaluation

In this section, after a brief description of the processing time required by the tool, we report on an experimental evaluation we made with a significant number of experts in the field of Transportation and cycling mobility. Finally, we report on the flexibility of the BikeScience libraries to work with different kinds of data and develop different kinds of applications and tools.

6.1 Performance

To be able to use BikeScience in interactive applications, we took special care with performance, leveraging the speed of the underlying data science libraries we used. In a laptop with Intel Core i7-8550U processor with 16 GB of memory and 8 MB cache, the system took 140 s to load the data for the 8 years of data (8.45 million trips) and 1.7 s to load a single month (March/2019). This load operation is typically performed at system startup, before the interaction with the user. Once the data is loaded into the system, users can interactively perform different kinds of analysis; for example, computing all mobility flows took only 11 s for the 8 years and 3.7 s for the latest month. Then, computing the most relevant tier took 16 s for the whole period and 0.4 s for the latest month.

6.2 Assessment with expert users

To evaluate the perception from real users, we conducted an assessment with 26 expert users: 11 from different departments of a City Traffic Engineering Authority, four from Non-Governmental Organizations (NGOs) related to traffic safety and promotion of bicycle use, two from a data analysis startup, three analysts from a dockless bike-sharing company, and six analysts from a dock-based bike-sharing company. The expert users were presented with the BikeScience tool in an interactive session showing the main functionalities and analyses presented in this paper. Then, we applied a questionnaire based on the Technology Acceptance Model (TAM) (Davis and Venkatesh 1996), a widely used model to assess how users perceive the *usefulness*, *ease of use*, and the *desire to adopt* a new technology. Table 7 shows the TAM questionnaire. It is composed of 10 items: four statements related to usefulness (U1–U4), four related to ease of use (E5–E8), and two related to behavioral intention to use (I9 and I10). The italic words are the main concepts addressed by each item. We used a seven-point *Likert* scale to gather the participants' degree of agreement with each statement, from (1) totally disagree to (7) totally agree.

As depicted in Fig. 14, most participants rated BikeScience well in the questionnaire, deeming the tool as easy to use (E5–E8), useful (U1–U4), and showing an intention to use it in the future (I9 and I10). Almost all responses to the TAM items were rated positively as *somehow agree*, *agree*, or *totally agree*. Also, we asked the participants their opinion about the tool. All participants deemed BikeScience as useful for transportation planning and for the analysis of bike-

Table 7 TAM statements

Perceived usefulness (PU)

U1 Using the BikeScience tool improves my *performance* in analyzing this kind of data compared to what I have today

U2 Using BikeScience would increase my *productivity* in this kind of analysis

U3 Using BikeScience would improve my *effectiveness* in this kind of analysis

U4 Overall, I find BikeScience *useful* to perform this kind of task

Perceived ease of use (PE)

E5 *Learning* to use the BikeScience tool seems easy to me

E6 Interacting with BikeScience does not require a lot of *mental effort*

E7 Overall, I find BikeScience *easy to use*

E8 I find it easy to rely on BikeScience *to get relevant information* on urban mobility with shared bikes

Behavioral intention of use (BI)

I9 Assuming I had access to BikeScience, I *intend* to use it for this kind of analysis

I10 Given that I had access to BikeScience, I *predict* that I would use it for this kind of analysis
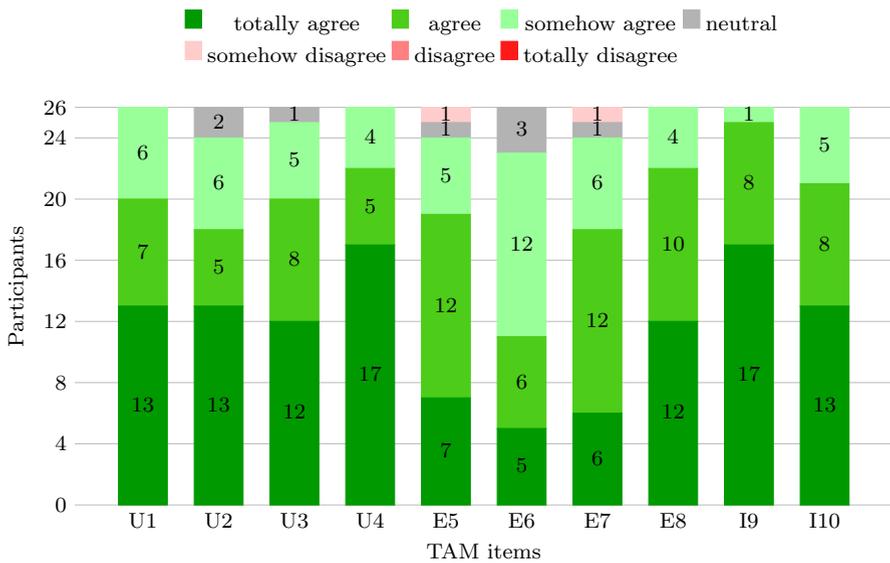
Fig. 14 Answers to the TAM assessment

sharing systems. Some of them suggested the inclusion of other data sources to improve the analysis, such as traffic accidents and OD survey data, which has been recently carried out with little effort given the high modularity of the BikeScience libraries and the flexibility of the tool.

According to the feedback from the experts participating in the study, the tool would support their existing workflow providing more concrete evidence on which

existing cycling lanes (1) should be prioritized for maintenance or (2) could eventually be discontinued; also, (3) validating the existing cycling infrastructure expansion plan, and (4) prioritizing specific construction works when implementing the expansion plan.

6.3 Flexibility case study

To demonstrate the flexibility of the BikeScience analysis libraries, we conducted a case study based on a request made by Transportation Engineers that participated in the assessment described before.

The Engineers mentioned that the analysis based on OD survey data could bring insights on how the bicycle and other means of transport are now used and could be used in the future. They requested that we implemented a visualization of significant urban flows of short car trips (< 4 km) that cover a path with little inclination ($< 4\%$), which could be potential candidates for a modal change from car to bicycle. They also mentioned that instead of having this analysis implemented as a set of Jupyter notebooks, it would be convenient to have a web-system that would not require any software installation in the local computer.

To test the flexibility of the BikeScience libraries for these two new requirements, we asked another member of our research group, who had no previous experience with BikeScience, to develop an interactive web application to perform the required analysis. He was able to implement the analysis requested by the expert users in a few hours of work. The implementation of an interactive web-based version required a little more work because he needed to learn how to use a few new libraries for web development. The system was then concluded in about 1 week of work. The resulting interactive web-based visualization was considered by the transportation engineers as strong evidence to support their planning decision-making workflow.

Then, we could conclude that the provided tool (distributed as a set of Jupyter notebooks) was considered highly useful by Transport Planners that do not have a programming background. Also, the Python libraries, which serve as the back-end for the tool, are highly flexible and can be used by programmers to quickly develop new functionalities and new kinds of interfaces when required.

7 Conclusions

In this paper, we introduced a method and associated open source tool capable of analyzing millions of bicycle trips from a city to derive relevant information about cycling mobility patterns. The method can be useful for urban transportation planners, BSS operators, and urban mobility researchers and fulfills the seven requirements listed in Sect. 3. It was validated with an analysis of over 8 million trips during the entire history of BSS in the Boston area. In particular, the procedure for abstracting flows at different levels can be a powerful aid for public planners to make decisions on infrastructure investments, educational campaigns, and law enforcement monitoring. For example, the analysis for Boston showed that the government

could take care of half of the bike-sharing trips by investing in only 3% of the flows; the BikeScience tool displays graphically which of these top flows are not well served by proper cycling infrastructure. The method can also be useful for businesses offering bike-related services and for bike-sharing system providers. We conducted an assessment of BikeScience with expert users following the TAM procedure. The results demonstrated that experts consider the method and the tool very useful, relatively easy to use, and they intend to adopt the tool in the near future.

After demonstrating the tool in front of the São Paulo Transportation Authority, they decided to incorporate BikeScience into their daily toolset as an aid for making decisions on planning the cycling infrastructure of the city. In particular, they will use the tool to decide the next steps in creating new segregated cycle tracks and painted bike lanes.

A limitation of our work is the fact that we are using bike-sharing trips as a proxy to the overall use of bicycles in the urban space. In many cities, however, bike-sharing represents only a fraction of the total cycling trips. It might be the case that cyclists who have their own bikes follow different mobility patterns. This issue could be mitigated in the future by using image processing on traffic videos to detect bicycle flows. The difficulty with this approach is that to have significant coverage of a city, this would require analyzing hundreds, or even thousands, of video cameras in real-time to detect and count bike flows. To the best of our knowledge, this has not yet been done at a scale that would allow any city-wide analysis of mobility flows. An alternative would be to use the location service of smartphones to detect cycling trips (e.g., based on speed). This would require tracking the location of millions of people, which has strong privacy implications but, in fact, is already performed by cell phone carriers and mobile OS vendors. We also observe that the number of trips between stations depends on the availability of bicycles on originating stations and docking spots on destination stations (Médard de Chardon et al. 2016). Thus, rebalancing strategies to meet high user demand in stations close to transportation hubs during rush hours may impact the distribution of trips between stations. Notwithstanding this fact, our method reflects the actual number of trips between all stations in the system.

Finally, the user study we conducted only shows that the users intend to adopt our technology and believe that it will be useful for their daily work. To demonstrate that it is indeed useful for the planning of new cycling infrastructure and to assess the quality of this plan, we need to wait 2 or 3 years before this assessment can be conducted.

As future work, we intend to continue developing our analysis method, including capabilities for comparing the patterns of multiple cities and correlating mobility flows with city socioeconomic and geographical data for flow prediction. A promising avenue for future research would be to extend our method with artificial intelligence, machine learning, and optimization capabilities. Given a certain budget, the method would aid urban planners in deciding where to make investments in the cycling infrastructure. For bike-sharing operators, it could be valuable to plug our technology into a real-time dashboard that would use both long-term and short-term historical data to not only visualize the current flows in the city but also predict the distribution of bikes in 1, 2 or 3 h in the future to help operators manage the system.

We also plan to offer our system to administrative authorities in cities that host a BSS to produce evidence to support policy-making. Currently, very few cities in the world adopt *scientific evidence-based policy-making* (Gray 2004). The analysis of the large amounts of data produced by cities with scientific rigor could provide a solid ground for the design of more effective public policies.

Compliance with ethical standards

Conflict of interest The authors state that there is no conflict of interest.

Source code and materials All the software described in this paper is available as open source at https://gitlab.com/interscity/bike-science.

References

Bao J, He T, Ruan S, Li Y, Zheng Y (2017) Planning bike lanes based on sharing-bikes' trajectories. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, p 1377-1386

Beecham R, Wood J (2014) Exploring gendered cycling behaviours within a large-scale behavioural dataset. Transportation Planning and Technology 37(1):83–97

Beecham R, Wood J, Bowerman A (2014) Studying commuting behaviours using collaborative visual analytics. Computers, Environment and Urban Systems 47:5–15

Birch CP, Oom SP, Beecham JA (2007) Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. Ecological Modelling 206(3):347–359

Boston Transportation Department (2017) 2017 BOSTON BICYCLE COUNTS Measuring how many people ride bikes in the City. Tech. rep., City of Boston, https://www.boston.gov/sites/default/files/document-file-02-2018/2017_boston_bike_counts_report.pdf

Bushell MA, Poole BW, Zegeer CV, Rodriguez DA (2013) Costs for Pedestrian and Bicyclist Infrastructure Improvements - A Resource for Researchers, Engineers, Planners, and the General Public. Tech. rep., UNC Highway Safety Research Center

Cambridge (2014) Cycle Tracks: A Technical Review of Safety, Design, and Research. Tech. rep., City of Cambridge

Chen Y, Zha Y, Wang D, Li H, Bi G (2020a) Optimal pricing strategy of a bike-sharing firm in the presence of customers with convenience perceptions. Journal of Cleaner Production 253:119905

Chen Z, van Lierop D, Ettema D (2020b) Exploring Dockless Bikeshare Usage: A Case Study of Beijing, China. Sustainability 12(3):1238, 18 pages

Cohen J (1977) Statistical Power Analysis for the Behavioral Sciences. Academic Press, Orlando, FL

Coleman H, Mizenko K (2018) Pedestrian and bicyclist data analysis. Tech. Rep. "DOT HS 812 205", NHTSA's Office of Behavioral Safety Research

Corcoran J, Li T, Rohde D, Charles-Edwards E, Mateo-Babiano D (2014) Spatio-temporal patterns of a Public Bicycle Sharing Program: The effect of weather and calendar events. Journal of Transport Geography 41:292–305

Davis FD, Venkatesh V (1996) A critical assessment of potential measurement biases in the technology acceptance model: three experiments. International Journal of Human-Computer Studies 45(1):19–45

Demaio P (2009) Bike-sharing: History, Impacts, Models of Provision, and Future. Journal of Public Transportation 12(4):45–49

Duarte F (2016) Disassembling bike-sharing systems: Surveillance, advertising, and the social inequalities of a global technological assemblage. Journal of Urban Technology 23(2):103–115

Duran-Rodas D, Chaniotakis E, Antoniou C (2019) Built environment factors affecting bike sharing ridership: Data-driven approach for multiple cities. Transportation Research Part A: Policy and Practice 2673(12):55–68

Faghih-Imani A, Eluru N, El-Geneidy AM, Rabbat M, Haq U (2014) How land-use and urban form impact bicycle flows: Evidence from the bicycle-sharing system (BIXI) in Montreal. Journal of Transport Geography 41:306–314

Faghih-Imani A, Hampshire R, Marla L, Eluru N (2017) An empirical analysis of bike sharing usage and rebalancing: Evidence from Barcelona and Seville. Transportation Research Part A: Policy and Practice 97:177–191

Fishman E (2016) Bikeshare: A Review of Recent Literature. Transport Reviews 36(1):92–113

Fishman E (2020) Bike Share. Routledge, Taylor & Francis

Flater D (2011) Understanding geodesic buffering. ArcUser Esri, Winter, pp 33–36

Frade I, Ribeiro A (2014) Bicycle Sharing Systems Demand. Procedia - Social and Behavioral Sciences 111:518–527

Garrard J, Handy S, Dill J (2012) Women and Cycling. In: City Cycling, MIT Press, pp 211–234

Gray JAM (2004) Evidence based policy making. BMJ 329(7473):988–989

Gu T, Kim I, Currie G (2019) To be or not to be dockless: Empirical analysis of dockless bikeshare development in China. Transportation Research Part A: Policy and Practice 119:122–147

Handy SL, Xing Y, Buehler TJ (2010) Factors associated with bicycle ownership and use: A study of six small U.S. cities. Transportation 37(6):967–985

He T, Bao J, Ruan S, Li R, Li Y, He H, Zheng Y (2020) Interactive bike lane planning using sharing bikes' trajectories. IEEE Transactions on Knowledge and Data Engineering 32:1529–1542

Hoevenaar-Blom MP, Wendel-Vos GW, Spijkerman AM, Kromhout D, Verschuren W (2011) Cycling and sports, but not walking, are associated with 10-year cardiovascular disease incidence: The MORGEN Study. European Journal of Preventive Cardiology 18(1):41–47

Hong Z, Mittal A, Mahmassani HS (2016) Effect of Bicycle-sharing on Public Transport Accessibility: Application to Chicago Divvy Bicycle-sharing System. In: Transportation Research Board, 95th Annual Meeting, pp 1–19

Huet N (2020) Chain reaction: commuters and cities embrace cycling in COVID-19 era. Tech. rep., Euronews. https://www.euronews.com/2020/05/12/chain-reaction-commuters-and-cities-embrace-cycling-in-covid-19-era

LAB (2018) The Bicycle Friendly America Award Database. Tech. rep., The League of American Byciclists, www.bikeleague.org/sites/default/files/bfareportcards/BFC_Fall_2017_ReportCard_Boston_MA.pdf

Lusk AC, Furth PG, Morency P, Miranda-Moreno LF, Willett WC, Dennerlein JT (2011) Risk of injury for bicycling on cycle tracks versus in the street. Injury Prevention 17(2):131–135

Maldonado-Hinarejos R, Sivakumar A, Polak JW (2014) Exploring the role of individual attitudes and perceptions in predicting the demand for cycling: a hybrid choice modelling approach. Transportation 41(6):1287–1304

Margie Peden RS, David Sleet DM, Adnan A Hyder EJ, Mathers C (2004) World report on road traffic injury prevention. Tech. rep., World Health Organization, Geneva

Médard de Chardon C, Caruso G, Thomas I (2016) Bike-share rebalancing strategies, patterns, and purpose. Journal of Transport Geography 55:22–39

Médard de Chardon C, Caruso G, Thomas I (2017) Bicycle sharing system 'success' determinants. Transportation Research Part A: Policy and Practice 100:202–214

NHTS (2018) National Household Travel Survey. https://nhts.ornl.gov/vehicle-trips, accessed: 2018-11-20

O'Brien O, Cheshire J, Batty M (2014) Mining bicycle sharing data for generating insights into sustainable transport systems. Journal of Transport Geography 34:262–273

Oja P, Titze S, Bauman A, de Geus B, Krenn P, Reger-Nash B, Kohlberger T (2011) Health benefits of cycling: A systematic review. Scandinavian Journal of Medicine and Science in Sports 21(4):496–509

Openshaw S, Taylor PJ (1979) A million or so correlation coefficients: three experiments on the modifiable areal unit problem. Pion, London, pp 127–144

Pedroso FE, Angriman F, Bellows AL, Taylor K (2016) Bicycle Use and Cyclist Safety Following Boston's Bicycle. Public Health 106:2171–2177

Ploeger J, Oldenziel R (2020) The sociotechnical roots of smart mobility: Bike sharing since 1965. Journal of Transport History 41(2):134–159

PSI (2019) Bike Sharing Market Research Report: By Type, Geographical Outlook - Industry Analysis and Growth Forecast to 2025. Tech. rep., Prescient & Strategic Intelligence Private Ltd

Raux C, Zoubir A, Geyik M (2017) Who are bike sharing schemes members and do they travel differently? The case of Lyon's "Velo'v" scheme. Transportation Research Part A: Policy and Practice 106:350–363

Ricci M (2015) Bike sharing: A review of evidence on impacts and processes of implementation and operation. Research in Transportation Business and Management 15:28–38

Richter F (2018) The Global Rise of Bike-Sharing. Tech. rep., Statista, https://www.statista.com/chart/13483/bike-sharing-programs

Sælensminde K (2004) Cost-benefit analyses of walking and cycling track networks taking into account insecurity, health effects and external costs of motorized traffic. Transportation Research Part A: Policy and Practice 38(8):593–606

Sarkar A, Lathia N, Mascolo C (2015) Comparing cities' cycling patterns using online shared bicycle maps. Transportation 42(4):541–559

Shaheen SA, Lipman TE (2007) Reducing greenhouse emissions and fuel consumption: Sustainable approaches for surface transportation. IATSS Research 31(1):6–20

Shaheen SA, Guzman S, Zhang H (2010) Bikesharing in Europe, the Americas, and Asia: Past, present, and future. Transportation Research Record 2143(1):159–167

Shen Y, Zhang X, Zhao J (2018) Understanding the usage of dockless bike sharing in Singapore. International Journal of Sustainable Transportation 12(9):686–700

Shephard RJ (2008) Is active commuting the answer to population health? Sports Medicine 38(9):751–758

van Acker V, van Wee B, Witlox F (2010) When transport geography meets social psychology: Toward a conceptual model of travel behaviour. Transport Reviews 30(2):219–240

Wang X, Lindsey G, Schoner JE, Harrison A (2016) Modeling bike share station activity: Effects of nearby businesses and jobs on trips to and from stations. Journal of Urban Planning and Development 142(1):04015001, 9 pages

Weiss M, Dekker P, Moro A, Scholz H, Patel MK (2015) On the electrification of road transportation - A review of the environmental, economic, and social performance of electric two-wheelers. Transportation Research Part D: Transport and Environment 41:348–366

Wong D (2009) The modifiable areal unit problem (MAUP), The SAGE Handbook of Spatial Analysis, London, chap 7, pp 105–123

Wood J, Slingsby A, Dykes J (2011) Visualizing the dynamics of london's bicycle-hire scheme. Cartographica: The International Journal for Geographic Information and Geovisualization 46(4):239–251

Yanocha D, Mason J, Patlán M, Benicchio T, Alfred T, Laksmana U (2018) The Bikeshare Planning Guide. Tech. rep., Institute for Transportation and Development Policy (ITDP)

Zhang L, Zhang J, Duan ZY, Bryde D (2015) Sustainable bike-sharing systems: Characteristics and commonalities across cases in urban China. Journal of Cleaner Production 97:124–133

Zhang Y, Brussel MJ, Thomas T, van Maarseveen MF (2018) Mining bike-sharing travel behavior data: An investigation into trip chains and transition activities. Computers, Environment and Urban Systems 69:39–50

Zhao J, Deng W, Song Y (2014) Ridership and effectiveness of bikesharing: The effects of urban features and system characteristics on daily use and turnover rate of public bikes in China. Transport Policy 35:253–264

Zhao J, Wang J, Deng W (2015) Exploring bikesharing travel time and trip chain by gender and day of the week. Transportation Research Part C: Emerging Technologies 58:251–264

Zhou X (2015) Understanding spatiotemporal patterns of biking behavior by analyzing massive bike sharing data in Chicago. PLoS ONE 10(10):0137922, 20 pages