MASTERARBEIT / MASTER’S THESIS

Titel der Masterarbeit / Title of the Master‘s Thesis

„Bike Sharing Atlas: Visual Analysis of Bike-Sharing Networks and Urban Commuting Patterns Worldwide“

verfasst von / submitted by

Michael Oppermann, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Diplom-Ingenieur (Dipl.-Ing.)

Wien, 2017 / Vienna 2017

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 926
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Wirtschaftsinformatik
Betreut von / Supervisor:	Dipl.-Inf. Michael Sedlmair, PhD
Mitbetreut von / Co-Supervisor:	Univ.-Prof. Torsten Möller, PhD

Declaration of Authorship

I hereby declare that I have written this Master's Thesis independently, that I have completely specified the utilized sources and resources and that I have definitely marked all parts of the work - including tables, maps and figures - which belong to other works or to the internet, literally or extracted, by referencing the source as borrowed.

Vienna, 4th May, 2017

Michael Oppermann

Acknowledgements

First, I want to thank Michael Sedlmair and Torsten Möller for their support and encouragement. I am extremely grateful for the opportunity to have had them as supervisors and mentors. Many discussions and their excellent feedback sparked my interest in doing further research in this area.

I also want to thank all members of the Visualization and Data Analysis research group for great discussions and their valuable feedback.

Furthermore, I would like to thank my family and all my friends for their patience and support.

Publication & Media

Journal Paper

Portions of this thesis have been submitted as a paper for publication in the *Special Issue on Visualization* of the *International Journal of Transportation* in April 2017. Thus, any use of “we” in this thesis refers to Michael Oppermann, Michael Sedlmair, and Torsten Möller.

Presentation

Michael Oppermann will present this work under the title ’A Global Perspective on Bike-Sharing Networks And Urban Commuting Patterns’ at the *8th International Visualization in Transportation Symposium* in Washington D.C. in July 2017.

Media

The Bike Sharing Atlas has attracted public interest and was also featured in the media:

1. Futurezone technology news, ’Bike Sharing Atlas zeigt, wo die Mieträder stehen’ [1]
2. uni:view, ’Mit dem Radl von Stadt zu Stadt’ [2]

Kurzfassung

In dieser Arbeit präsentieren wir die interaktive, web-basierte Visualisierung Bike Sharing Atlas (bikesharingatlas.org). Unser entwickeltes System erlaubt die explorative Datenanalyse von mehr als 468 Bike-Sharing Netzwerken weltweit. Über einen Zeitraum von 17 Monaten haben wir Live-Daten hunderter Bike-Sharing Netzwerke aufgezeichnet, mit weiteren Informationen angereichert und eine umfassende Datenbank erstellt. Mehrere verknüpfte Ansichten und eine neue Interaktionstechnik für Liniendiagramme erlauben Benutzern die Identifizierung von Kapazitätsengpässen bei Stationen, Vergleiche von Netzwerkcharakteristiken, oder beispielsweise die Analyse von Pendlerströmen in hunderten Städten. Das übergeordnete Ziel dieser Arbeit ist, zu illustrieren welches Potenzial die visuelle Analyse für die Exploration von verteilten, heterogenen Daten von Smart Cities bietet. Basierend auf Gesprächen und Evaluierungen mit Personen verschiedener Zielgruppen, präsentieren wir vier exemplarische Einsatzmöglichkeiten. Diese Szenarien demonstrieren das Potenzial unseres visuellen Ansatzes für ein besseres Verständnis von Bike-Sharing und urbanen Pendlerbewegungen in einem globalen Kontext.

Abstract

In this paper, we introduce an interactive visualization system, *bikesharingatlas.org*, that supports the explorative data analysis of more than 468 bike-sharing networks worldwide. Being increasingly digitized, these networks nowadays produce data that can reveal interesting insights, not only into patterns of bicycle usage but also underlying spatio-temporal dynamics of a city. We recorded this data from several hundred networks worldwide, over a period of 17 months, and made it publicly accessible through a common web platform. The application leverages a multi-coordinated view approach and innovative interaction techniques can help, for instance, to expose capacity bottlenecks, commuting patterns, and other network characteristics. Our broader goal is to illustrate how visual analysis can be used for exploring distributed, heterogeneous data from smart cities. Based on our collaboration with different target users, we present a set of four usage scenarios that show the potential of our approach to understanding bike-sharing and urban commuting behaviors worldwide.

Contents

CHAPTER 1

Introduction

1.1 Motivation

The majority of the world's population is living in urban areas, and this proportion continues to grow. An increase in efficiency is needed for cities to function, and sustainable infrastructures will be essential to accommodate larger numbers of people. As digitization has become an integral part of our life, massive amounts of data from a variety of sources are generated continuously in cities worldwide. Leveraging this data intelligently offers great potential towards smarter and more efficient cities.

Although having these immense datasets at our fingertips, we often lag behind in supporting people to intelligently leverage the huge amount of data that is produced daily. A city planner might want to identify and better understand commuting patterns in order to develop more robust and cohesive transportation infrastructures. A local politician wants to make information more accessible and communicate decisions in a clear and understandable way. A journalist might want to write an article about sustainable mobility and substantiate various hypotheses with real data. A sociologist wants to study local effects of job density and residential segregation on society, or wants to perform other data-intensive tasks like cross-country comparisons of urbanization.

However, currently, most of the data comes in machine-readable form only and hence is hard to access for people without sophisticated computational and statistical skills.

In this thesis, our goal is to illustrate how interactive visualization can help to open the data that is produced in smart cities to a wider audience. We believe that interactive visualization can help us to engage users, and to interactively explore and understand collected data from smart city sensors, in order to make life more comfortable, safer and sustainable.

Towards this goal, we take public bike-sharing systems as an example and show how visualization can help to better leverage the data produced by these systems. Bike-sharing systems have been established as permanent components in urban passenger transport since 1996 [3]. Users can rent a bicycle at any bike-sharing station, drive to their destination and then return it again at a nearby station. Due to the fast and easy access and the high station density in many cities, bike-sharing is becoming more and more popular for inner-city transportation. Being increasingly digitized, these systems nowadays produce data that can reveal interesting insights, not only into patterns of bicycle usage, but also underlying spatio-temporal dynamics of a city, as Froehlich et al. [4] and Wood et al. [5] pointed out. We gathered this data from several hundred cities worldwide and made it publicly accessible through a common web platform. For this purpose, we have implemented an interactive visualization system with multiple linked views that allow an exploratory data analysis for users with a wide range of expertise.

1.2 Methodological Approach

In November 2014 we conducted a brainstorming session with the operator of a bike-sharing network in Vienna. We discussed several challenges they encounter during setup and operation of a cost-oriented bike-sharing network and also talked about utilizing visual analysis. This meeting and the provided datasets motivated us to implement a first prototype that supports the interactive exploration and analysis of different aspects of an individual bike-sharing network. Follow-up conversations via email and in person showed us that the responsible people gathered a wealth of experience during the past decade of operation. While our tool can support and accelerate various analysis processes, it did not reveal entirely new insights. Interestingly, although bike-sharing is a global trend with opportunities and challenges that occur not just locally, we could observe that the operator and existing visualization tools for other cities take a rather narrow view. With this knowledge in mind, we concluded the pre-project stage and started to develop our idea of a global approach to understanding smart city data.

Our subsequent methodological design and approach is composed of three main steps: (1) data acquisition and preprocessing, (2) implementation, and (3) evaluation.

As a first step, we recorded continually, over a period of 17 months, distributed data from several hundred bike-sharing networks worldwide. We aggregated the data, combined it with other data sources and derived characteristic network metrics. In the course of an iterative design process we built a visualization tool with a variety of viewing options to offer multiple different perspectives on the data. Various design considerations were inspired and strongly shaped by our interactions with different target user groups: operators of a bike-sharing network, urban sociology researchers, public authorities, city planners, and the general public.

To illustrate the value of our proposed visualization approach, we provide four different usage scenarios:

1. how visualizing long-term recorded data can help to better understand commuting patterns in a city,

2. how the combination with other data sources can reveal interesting insights,
3. how our approach allows to compare systems world-wide through a shared global perspective,
4. and how the general public can benefit from recording and visualizing this data.

While our focus is on bike-sharing networks, we believe that—with the proliferation of globalization and mega cities—such data-driven and visual approaches will become increasingly important for other aspects of open urban data as well.

Each step of our methodological approach is described in the Chapters 3, 4 and 5. Finally, in Chapter 6, we reflect on the implications of our approach and discuss opportunities for other smart city initiatives.

1.3 Contributions

In summary, our main contributions are:

- We recorded and make openly available a repository of open bike-sharing data from 380 different cities over a period of 17 months and additionally data from 88 cities over a period of 4 months.
- We designed an online interactive visualization that makes the data accessible to users with a wide range of expertise.
- We describe a set of four usage scenarios to illustrate the benefits of our approach for different target users.

CHAPTER 2

Background and Related Work

2.1 Bike Sharing

Bike-sharing networks—established as permanent components in urban passenger transport since 1996 [3]—are increasingly pervading and influencing the way we commute in densely populated areas worldwide. New systems are popping up and existing ones are extended continuously. A bike-sharing network, by our definition, has a certain number of *stations* and often thousands of bikes circulating in the network. Users can rent a bicycle at any station, drive to their destination and return it at a station nearby. Every station is composed of docking spaces and has, therefore, a finite capacity. The number of bikes in each station, we call it *fill level*, is highly dynamic and also a significant factor for the functioning of a system. There is an increasing number of public bike sharing systems that do not rely on these docking stations and instead work with GPS-equipped bicycles that can be dropped off at any major crossroad or in a geo-fenced area. The renting and returning procedures work with smart cards or apps. Given the still rather small number of these *dockless* systems and their different characteristics, the main focus in this thesis lies on the data produced by the first-mentioned, *traditional* bike-sharing systems.

Correctly deployed, these systems can offer a wide range of benefits. Commuters do not have to buy their own bicycles in the first place and, second, they do not have to bother with security, maintenance, and shelter. These systems benefit not only the users but also the city as a whole. They can relieve pressure from overcrowded routes in the public transportation system and can act as a complement to traditional urban transport options. For example, to cover the last mile between the subway station and the workplace. The installation and maintenance are relatively cost-efficient compared to other transportation options, thus, bike sharing programs provide a good opportunity to promote sustainable mobility and to reduce car usage in the city centers.

Despite the advantages offered by these networks, the initial setup and the operation can lead to many challenges and uncertainties for the responsible authorities. The primary objective, shared between all operators and city authorities, is a *self-balanced* bike-sharing network. In this scenario, the number of available bikes and free docks are distributed in a way that users can rent bicycles wherever and whenever they want. In practice, stations get full or empty and operators have to redistribute bikes manually between the stations in order to guarantee the functioning of a network. This process can be very complex and waste enormous resources. There are numerous papers addressing this rebalancing issue, for example, by suggesting more efficient routes for the manual redistribution or a more sophisticated positioning of docking stations (see Section 2.2). In this context, questions arise, such as: How many stations and docking spaces are needed? What average distance between stations is adequate? What location generates usage throughout the day?

Being increasingly digitized, the systems produce large amounts of data that can reveal interesting insights and can help to answer these types of questions. Subsequently, this can result in an improvement of bike-sharing programs in the future. By collecting data that is produced by these systems continuously (see Chapter 3) and by making it publicly accessible and explorable in an interactive visualization system (see Chapter 4) we want to make a further step in this direction and show the potential of a data-driven approach to understanding bike-sharing networks and smart cities in general.

2.2 Related Work

Bike-sharing data has been used in different contexts before. Here, we review this previous work with the goal to contextualize our work and to provide the interested reader with further pointers into the visualization and data analysis literature.

2.2.1 Statistical Analyses of Bike Sharing Networks

The most prominent analyses of bike-sharing data this far have been statistical analyses. The goal of these analyses were, for instance, to help system operators to improve the location of stations [6, 7], or re-balance bikes between stations [8, 9, 10]. Lin et al. proposed a model to determine the optimal number of stations to cover a specific area and to ensure that the network can handle the traffic effectively [11]. Guenther et al. developed models to forecast future bicycle migration trends in order to predict station fill levels [12] and Borgnat et al. used statistical modeling to analyze dynamics of movements within Lyon's bike-sharing scheme [13]. While these approaches look for an algorithmic solution for a clearly defined task, our goal is different in that we want to make the data accessible, visible, and explorable to a wide range of potential users.

2.2.2 Visual Analysis of Bike Sharing Networks

Towards our goal, to open the data and to make it accessible to a wider audience, we took the approach of interactive visualization, which is well-known to support such exploratory endeavors [14, 15]. It provides an essential way to support the user in a hypothesis generation and decision-making process.

There exist already some efforts that focus on the visual analysis of individual bike-sharing networks. Studying the spatial distribution of journeys and peak-time behaviors in London [5, 16], analyzing the system dynamics over a 10-month long period in New York City [17] or investigating the effect of weather and calendar events on the spatio-temporal dynamics of Brisbane's bike-sharing network [18]. In addition to these papers, data scientists and students created numerous web applications and visualizations that also deal with the visual analysis of single bike-sharing networks. Wellington created a Voronoi tessellation of New York City to analyze bike-share riders [19]. Alberts mapped *Citi Bike* trips that were recorded over a two-day period [20] in New York City. Bostonography provides a tool to interactively explore the impact of demographic and weather factors on bike-share ridership in Boston [21]. Chiraphadhanakul implemented a well-designed interface for the exploration of bike-share data in the San Francisco Bay Area [22] and Jacobsen published an online tool that allows the comparison of Divvy riders versus public transit riders regarding travel time [23].

In terms of analyzing multiple cities, Austwick et al. [24] used statistics and visualization techniques to compare five different cities. Bargar et al. [25] proposed an application for comparing usage patterns between different bike-sharing programs from up to three different cities and Nagel et al. [26] exhibited, in a public gallery space, several visualizations for casually analyzing three different bike-sharing networks. Most closely related to our approach are O'Brien [27] and Meddin et al. [28], who both used map-based tools to show the locations of bike-sharing networks worldwide. Although these tools provide an overview of systems, they are based on very homogeneous data sources and do not take a long-term and contextual perspective, as we do. O'Brien et al. [29] proposed a classification of 38 systems based on spatio-temporal characteristics and demonstrated thereby the opportunities of a higher-level view. In contrast to this work, which is a static analysis and a discussion of various insights, our goal was an interactive visualization tool that allows for dynamic and interactive exploration of the underlying data.

2.2.3 Visual Analysis of Urban Data

In a broader context, there are several approaches that use other data sources in conjunction with visualizations to solve problems that are abstractly similar to the one of bike-sharing networks. These works use other smart city sensors, such as mobile phone data [30], public transport data [31] or social media data [30] to explore opportunities that arise from this type of data-driven analysis for smart cities. Ferreira et al. [32] used taxi trip data and Miranda et al. [33] utilized meta-data of Flickr images to visually explore and understand behavioral patterns in New York City. With these tools we share the goal to make the data accessible and to provide different levels of analysis. However, all these approaches take a local perspective on one city and are often limited to a specific target group, while our goal is to investigate how to expand such approaches to a more global level and for a broad range of users.

CHAPTER 3

Data Acquisition and Preprocessing

We gathered data from several hundred cities worldwide over a period of 17 months and combined it with other types of information, such as elevation profiles and hourly weather records. In the following, we more closely describe this underlying data and the preprocessing steps, before we introduce the interactive visualization system build upon it in the next section.

3.1 Bike Sharing Data

Many operators make their collected data available to the public and contribute to various open data initiatives. A few cities share detailed historical data about completed trips (see Table 3.2) but the majority provide only information about current station fill levels. In our application we put the focus on this type of information, to cover as many cities as possible.

We started collecting data from 380 networks at the beginning and added new networks continuously. Our database is now composed of data from 468 networks in 45 different countries, with more than 21.500 stations. We gathered the data through api.citybik.es [34] or directly from the websites of the bike-sharing operators. Over a period of 17 months and for 380 networks we have logged the number of available bikes and empty docks for each station in a 15 min interval (see Table 3.2). Additionally, we further recorded this information from 88 other networks over a 4-month period. The usage fees and membership plans, for networks with more than 100 stations, were collected manually from the websites.

[1] The data is not publicly available but the operator provided the information for our research project.

	Network	Region	Producer	Stations
1	Santander Cycles	London, United Kingdom	PBSC	774
2	Citi Bike	New York City, United States	PBSC	511
3	Divvy	Chicago, United States	PBSC	475
4	Ecobici	Mexico City, Mexico	Clear Channel	444
5	Capital Bikeshare	Washington DC, United States	PBSC	357
6	Nice Ride Minnesota	Minneapolis, United States	PBSC	190
7	Hubway	Boston, United States	PBSC	155
8	Bay Area Bike Share	San Francisco Bay Area, United States	Motivate	70
3	*Citybike* [1]	*Vienna, Austria*	*JCDecaux*	*200*

Table 3.1: Bike-sharing systems that share detailed information about completed trips

internal_id	station_id	network	slots	empty_slots	free_bikes	timestamp
108535	30184e9e...	citybike-wien	27	16	11	2017-03-25 18:45:02
108536	a2599132...	citybike-wien	15	0	15	2017-03-25 18:45:02
108537	f5e847ac...	citybike-wien	26	19	7	2017-03-25 18:45:02
108538	0ab8a7d9...	opole-bike	15	7	8	2017-03-25 19:01:11
...						

Table 3.2: Excerpt of station timestamps that are recorded every 15 min

We examined the data quality periodically to analyze the trustiness of the used APIs. In general, all station fill-levels are published in real-time and are consistent with the information on the websites of the individual bike-sharing networks. Due to the large number of networks and involved companies, it happens occasionally that no or only incomplete data is sent. These problems occur more likely with small systems and because of the long-time recording and the data aggregation, they have only little effect on the overall results. The complete absence of networks on our web application constitutes a larger problem. Many operators either do not publish station fill levels or they restrict the access for automatic processing. For example, the current version supports only a few systems from Asia, due to access limitations [34]. The implementation of special scrapers for individual networks is beyond the scope of this thesis but our tool provides interfaces to easily connect additional networks in the future.

3.2 External Data Sources

From the very beginning, our goal was to combine the bike-sharing data with external data sources and to investigate its benefits. Therefore, as a first step, we enhanced the platform with hourly weather records, elevation profiles and population statistics for all 468 networks.

Restrictions in accessing historical weather data made it necessary to implement a logging system that stores live weather records in our database. Since December 15, 2015 we use the API service of *OpenWeatherMap* [2] to access current weather data for all cities in our database. In our web application we are using only temperature data at the moment but we additionally save the textual weather description (e.g. overcast clouds, light rain, mist, etc.), humidity, and windspeed, that can be used for further analyses and extensions of the system.

The elevation profiles (altitude of each station) were loaded once by using the *Google Maps Elevation API* [3]. This information allows users, for example, to investigate the impact of elevation differences on the fill levels. Additionally, we are using the *Google Maps Geocoding API* [4] for converting addresses into geographic coordinates (latitude/longitude), and *Mapbox* [5] as a tile provider and for calculating routes.

Our interviews with various experts revealed a great interest in the integration of demographic data. Although there are many open data libraries worldwide, it is still a very complicated and time-consuming procedure to collect this type of information, especially on a city level. For this reason, we have restricted our search to city populations. We implemented a separate tool that can search for cities on Google and automatically scrapes the data from the population widget on top of the page. Missing records were added manually with data from Wikipedia. This information is only used as a rough indicator for filtering and clustering purposes and, therefore, possible uncertainties can be tolerated.

3.3 Preprocessing

In total, we collected more than 830 million fill levels, that offer a large potential for various analysis and prediction tasks, while also confronting us with additional processing challenges. For this reason, we integrated multiple preprocessing steps to break down the database into smaller chunks that can be loaded during runtime. Our Python scripts regularly parse the table with all timestamps, aggregate the data and export the result as small CSV files. Currently, we distinguish three modes of aggregating station fill levels. We calculate the average utilization of every station for the whole time period, for all weekdays or only for weekends. This process can be easily extended, for instance, to analyze seasonal patterns. Similarly, the daily average temperature is computed by aggregating hourly weather records.

3.4 Network Characteristics

High-level network metrics were also generated in advance. In addition to the number of stations per network, their elevation profiles, and the average number of docks per

[2] https://openweathermap.org/API
[3] https://developers.google.com/maps/documentation/elevation
[4] https://developers.google.com/maps/documentation/javascript/geocoding
[5] https://mapbox.com

station, we calculated the following characteristics for filter, ranking and comparison purposes:

Average nearest neighbor distance. For all the networks we created a distance-matrix to calculate the average nearest neighbor distance. The distances between the stations (latitude/longitude pairs) were computed by using the Haversine formula.

Average number of stations within a 2 km radius reach. Evaluations with the bike-sharing operator revealed that the averaging of distances can lead to inaccuracies. Therefore, we added another metric that is also used by the operator for internal analyses. The 2 km radius around a station represents very well the distance that users are willing to drive on average.

Network activity. The missing data about completed trips (e.g. origin and destination) from most of the registered systems makes it difficult to quantitatively assess bike-sharing dynamics. The mere aggregation of fill levels does not give us any insights about temporal changes. For instance, how many trips were taken during the week compared to the weekend? We defined a new metric, the network activity, by counting the number of fill level changes for every station and normalized it to the network size. Due to the 15 min interval in our data retrieval, there is some natural uncertainty of this straight-forward measure for larger networks [35].

Maximum elevation difference between the highest and lowest station. This metric can be used to filter networks with large elevation differences to further investigate its impact. Table 3.3 shows ten networks with the largest elevation drops and Figure 3.6 gives an overview of elevation differences between the highest and lowest stations in all our networks.

3.5 Statistics

In this section we present various statistics and visualizations that illustrate the characteristics as well as commonalities and differences between the bike-sharing networks in our database.

Figure 3.1 shows an overview of the largest bike-sharing networks in our database. These networks are not confined to a specific region and thus exemplify the global trend toward this type of sustainable urban mobility solution. Due to access limitations and frequent fluctuations in recent years this ranking does not reflect the largest networks that currently exist.

As shown in Figure 3.2 (logarithmic scale), networks with more than 200 stations may be considered as outliers. The majority of bike-sharing networks consists of less than 50 stations and the median is at 12.

The scatter plot in Figure 3.3 shows the relationship between the number of stations per network and the average nearest neighbor distance between stations. We can observe that larger networks have a much higher density and neighboring stations can be reached

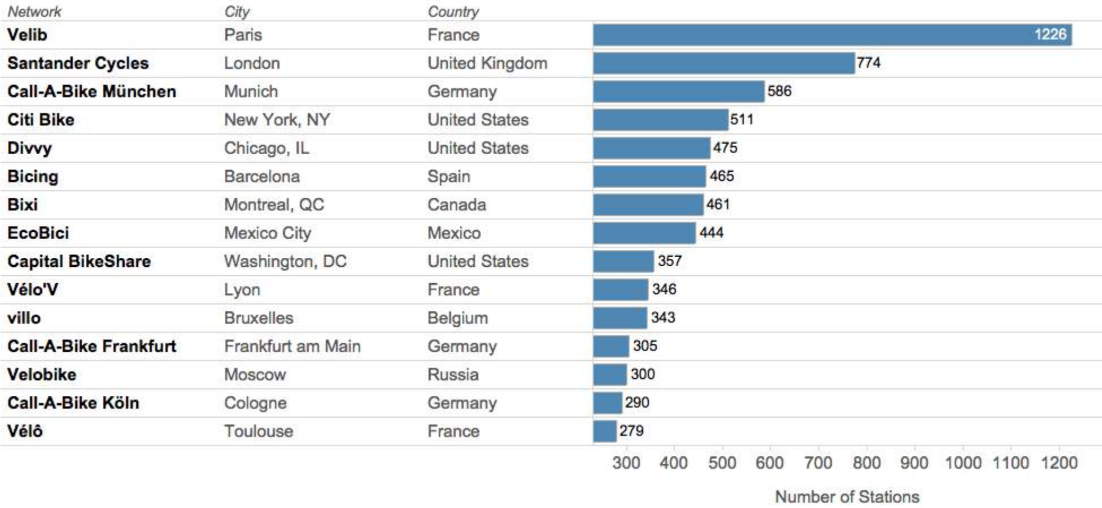


Figure 3.1: The 15 largest bike-sharing networks in our database

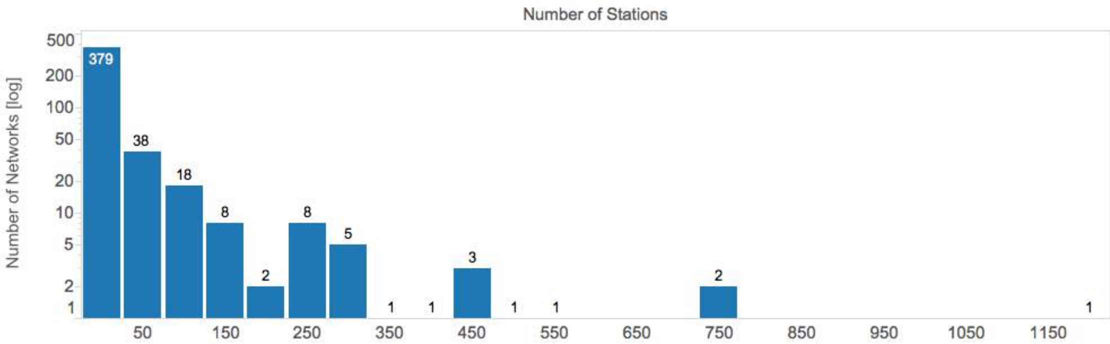


Figure 3.2: Histogram showing the distribution of number of stations per network

much faster. The average nearest neighbor distance is between 200 m and 900 m in most networks (see Figure 3.3) and the average number of stations within a 2 km radius reach is generally rather low, due to small networks sizes (see Figure 3.5). In the *Velib* network in Paris users can reach 120 stations on average within a 2 km reach. The walking distances to and from these stations in the city center have been reduced to a minimum. Regarding station density, Paris is followed by Barcelona, Mexico City and London.

Figure 3.7 (logarithmic scale) shows the population distribution of all cities in our database. Similar to the number of stations, the majority of networks is again at the lower and of the spectrum. Most of the bike-sharing networks are installed in cities with less than 500.000 inhabitants while there are also networks in mega cities such as London, Mexico City or New York City.

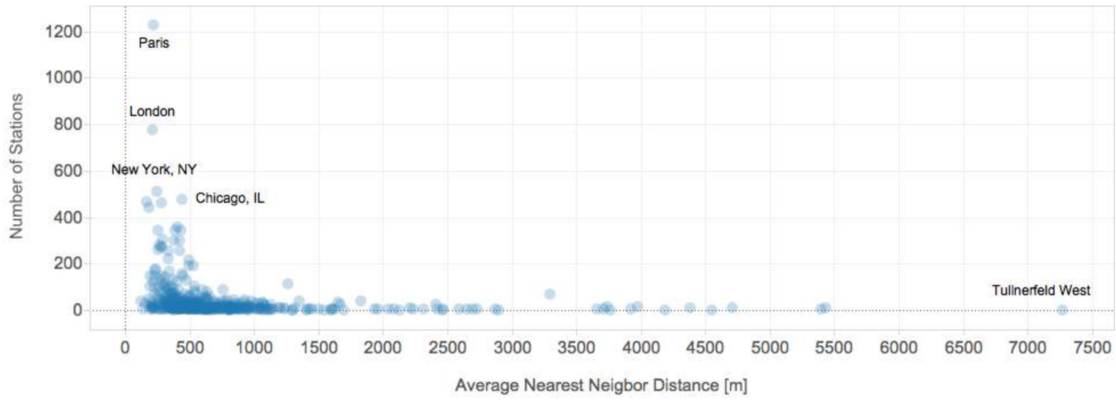


Figure 3.3: Number of stations vs. average nearest neighbor distance

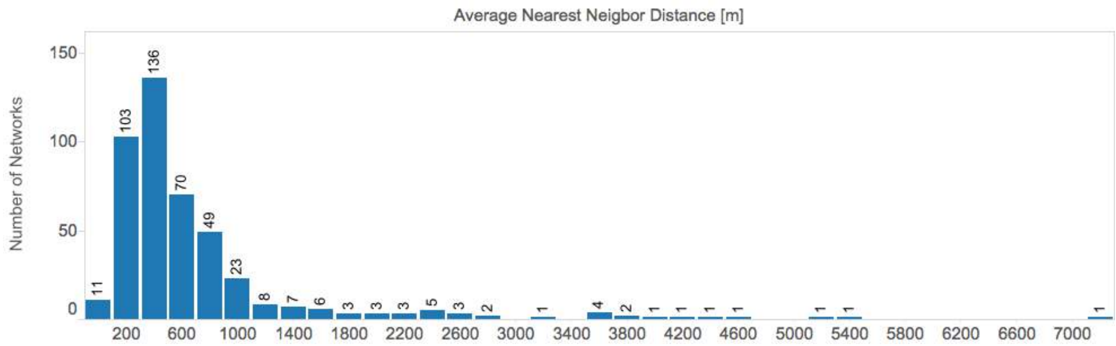


Figure 3.4: Histogram showing the distribution of the average nearest neighbor distance between bike-sharing stations

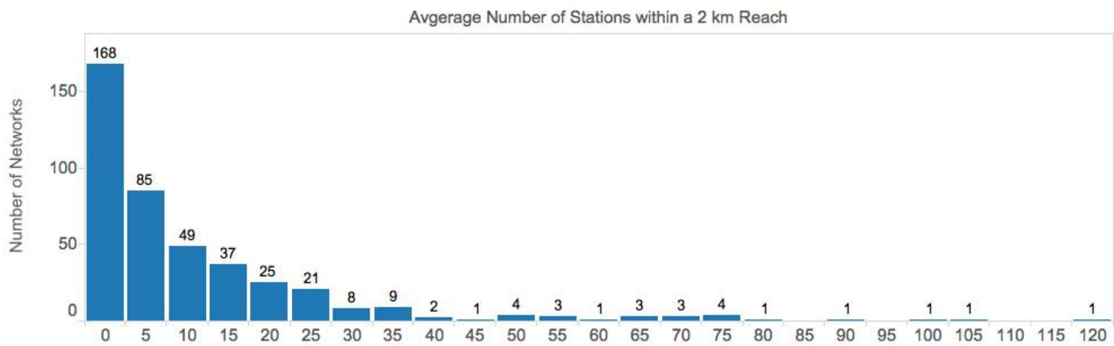


Figure 3.5: Histogram showing the distribution of the average number of stations within a 2 km radius reach

	Network	Region	Stations	Min. [m]	Max. [m]	Drop [m]
1	Parkinbici	Ischitella, Italy	12	1.42	566.16	564.74
2	Montana Valli dell'Ossola	Santa Maria Maggiore, Italy	6	271.01	830.81	559.79
3	Bike Santiago	Santiago, Chile	148	481.01	855.28	374.27
4	Terni	Terni, Italy	13	117.80	447.79	329.99
5	Enna	Enna, Italy	5	644.54	943.86	299.32
6	Youbike	New Taipei, Taiwan	343	1.18	267.38	266.20
7	e.motion	Rovereto, Italy	14	170.10	435.44	265.34
8	Mountain Rides	Ketchum / Sun Valley, USA	49	1610.01	1872.20	262.19
9	Call-A-Bike Stuttgart	Stuttgart, Germany	45	220.64	482.13	261.49
10	Leihradl	Traisen-Gölsental, Austria	8	328.52	583.01	254.50

Table 3.3: Ten networks with the largest elevation drop between the highest and lowest station

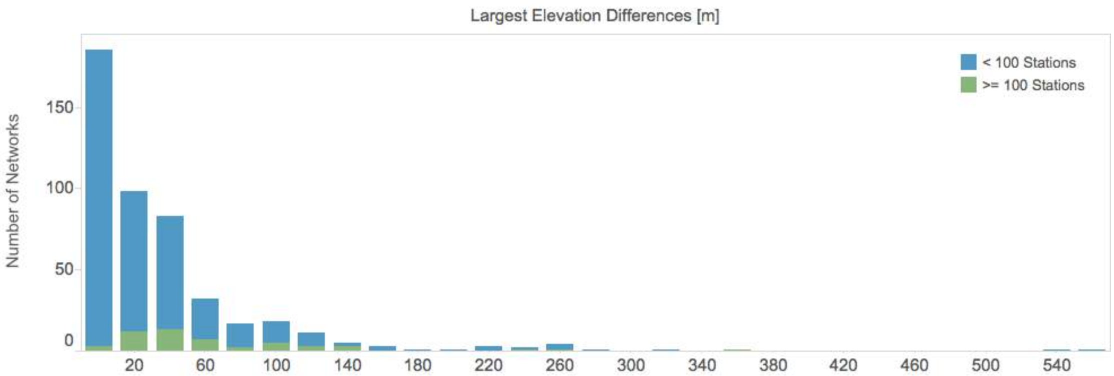


Figure 3.6: Histogram of the elevation difference between the highest and lowest station in each network

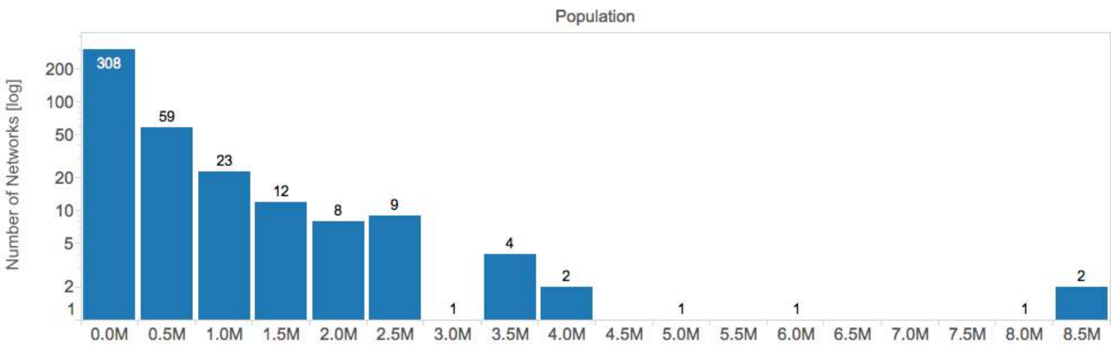


Figure 3.7: Histogram showing the distribution of population throughout the bike-sharing cities

CHAPTER 4

Bike Sharing Atlas

We have developed an interactive visualization system (bikesharingatlas.org) to help users with a wide range of expertise to understand and intelligently leverage data that is produced by public bike-sharing systems worldwide. This section contains multiple screenshots of the web tool and a video in the supplemental material provides further details on user interactions.

In the following, we give an overview of the system and subsequently we present four selected usage scenarios. In this context, we also describe in detail primary visual and interaction design choices we made during the implementation process.

4.1 System Overview

Global view: Initially, the homepage of the web-platform presents an interactive map showing the geographical locations of all bike-sharing networks. To ease getting started, we also display example cities including the city of the user's current location, and those that are particularly interesting because of the available data (see Figure 4.1). Two separate pages provide further high-level overviews of all networks, as shown in Figure 4.2 and 4.6. A small multiples view (Figure 4.2-b), for instance, provides a first impression of the size and density of the networks by visualizing them as individual vector maps. Sort functions, range sliders, and histograms allow users to explore and compare hundreds of networks in an interactive way.

Local view: The user can then select a network of a particular city and go one level deeper revealing detailed information about this network (and city). A tab navigation and multiple views provide different perspectives on the selected network and support the user in the exploration process. The currently available detail views include: (1) Interactive map with current fill levels of the network, as shown in Figure 4.4; (2) Route planner, Figure 5.5-c, and in Figure 4.5: (3) *Fill level analyzer* with historical data; (4)

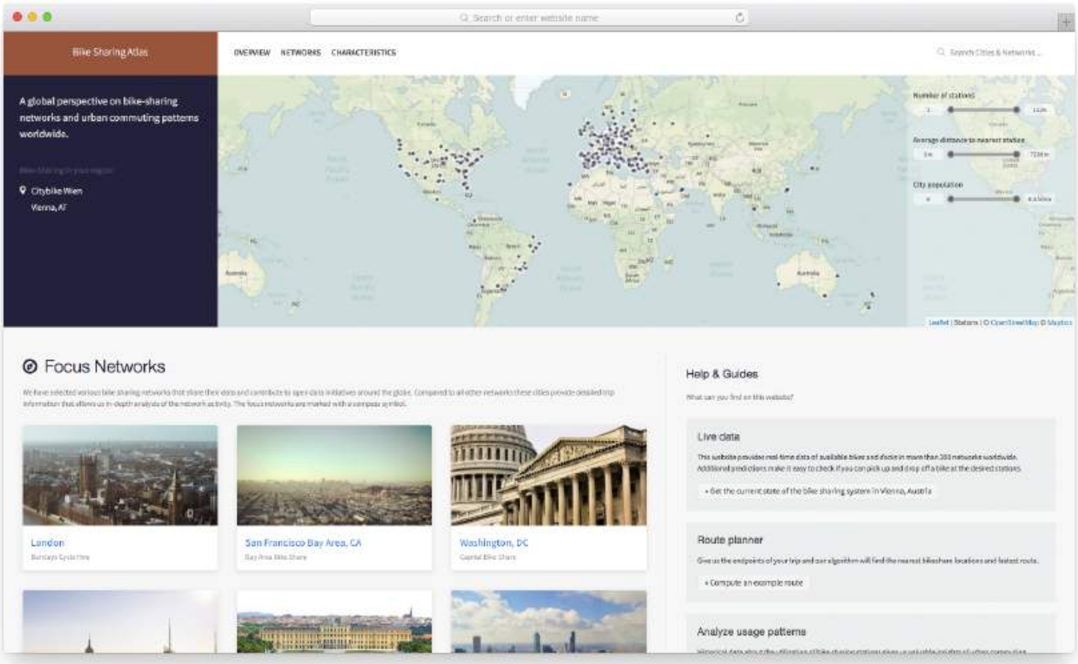


Figure 4.1: Screenshot of the homepage of *bikesharingatlas.org*

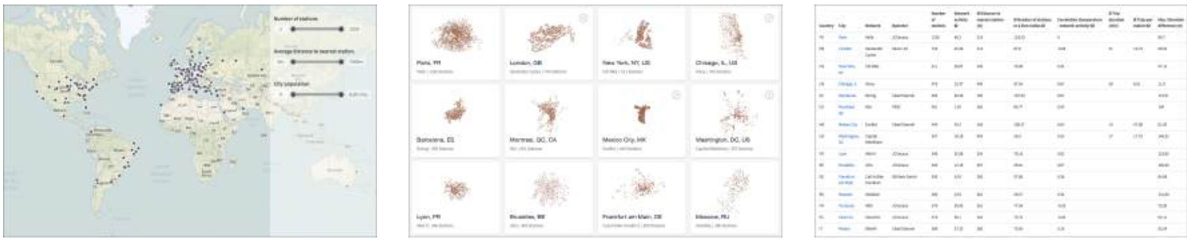
Time series chart with the network activity and the superimposed temperature profile; (5) Other information about the network and the city, such as bike-sharing pricing, detailed information on trips (if available), or additional weather and elevation data.

Multiple entry points, a search function, and a clear navigation structure provide an easy way to get from a global to a local view and vice versa. Moreover, we followed the idea of suggested interactivity [36], in the form of little preview videos and tooltips, to guide first-time users through the visualization system.

4.2 Implementation

Our system is implemented as a web-based tool that runs in every modern web browser and that adapts flexible to different desktop monitors. While we have used Python extensively for data-preprocessing, the actual web-interface is primarily based on HTML, CSS and JavaScript.

The collected and aggregated data is stored in a MySQL database and partly in small CSV files that can be loaded asynchronously as needed. We have decided to use MySQL because it is robust and it provides a straightforward way for representing our data



(a) Interactive filterable map provides a geographical overview of all networks

(b) Small multiples are good indicators of the size, structure and density of networks

(c) Table view with detailed statistical data of all networks

Figure 4.2: Screenshots of different overview components of the tool

model, and despite the fact that we would gain minor performance enhancements with NoSQL databases, such as MongoDB [1].

The REST API, the backend of the application, that serves as an interface between the MySQL database and the frontend is written in PHP and uses the micro-framework *FlightPHP* [2]. Most of the computations are performed in the preprocessing steps due to the vast amount of data (see Section 3.3 for more details). Remaining calculations that can be done during runtime are computed on the server. The frontend receives the prepared data and builds the graphical user interface. All the visualizations, interaction mechanisms, and event listeners are implemented in JavaScript. The advantage of this client-sever architecture is that we can operate multiple clients and different user interfaces with the same backend system. In addition to the Bike Sharing Atlas, we could implement, for example, a mobile app that builds on the same data and that can be easily connected to the server through our REST API.

Furthermore, we use the library *d3.js* [3] for visualizations, and *leaflet.js* [4] for integrating interactive maps. In contrast to other visualization libraries, d3.js is flexible enough to build unique visualizations with new interaction techniques while at the same time we don't have to rewrite basic functionalities from scratch. Leaflet is an open source JavaScript library for interactive web-based maps. The resulting mobile-friendly maps are similar those created by Google Maps but the library also allows us to integrate tile layers and plugins from many third-party providers. In our system, we are using *Mapbox* [5] as a mapping service for the image tiles and for calculating the routes in the bike-sharing route planner.

[1] https://mongodb.com
[2] http://flightphp.com
[3] https://d3js.org
[4] http://leafletjs.com
[5] https://mapbox.com

4.3 Visualization and Interaction Design Choices

The general goal of this tool is to help users with a wide range of expertise to understand and intelligently leverage data produced by a public bike-sharing system. To do so, we identified and followed several requirements in our design process that we further explain in the following sections.

4.3.1 Overviews and details

Conceptually, our user interface follows Shneiderman‘s venerable information-seeking mantra (“overview first, zoom and filter, details on demand”) [37]. This concept—one of the main principles of visualization designers nowadays—describes how data should be presented to users to be as effective as possible. Multiple views in our application show information about the entire dataset to provide an overview. For example, all the networks are displayed in an interactive map without showing any details but giving the user the ability to understand the data as a whole. Once the user can see the big picture, zooming, sorting, and filter mechanisms help to focus on a particular section of the data. Finally, once the user finds an interesting subset to look at, we provide further details on demand. In our visualization system, these details are usually different perspectives on individual bike-sharing networks (e.g. live station states, fill level analyzer, route planner, etc.). With the overview-detail design choice, we present the data in different levels of detail without visually overwhelming the user. Figure 4.3 shows the implemented components of the system.

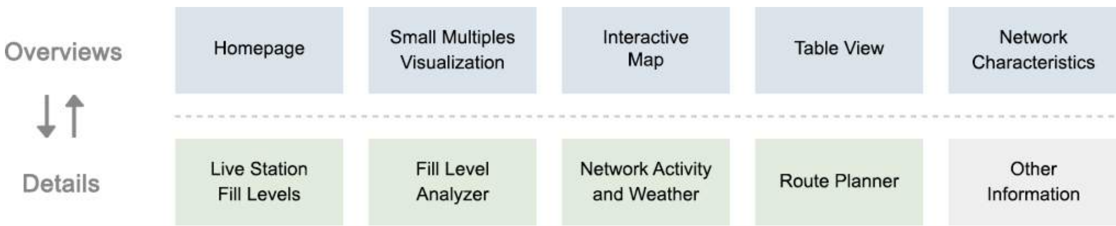


Figure 4.3: Overview and detail components.

4.3.2 Allow analysis-interested users to discover

While our platform is meant to be accessible to everyone and easy to use, at the same time we sought to encourage users to explore the data and foster serendipitous discoveries. Towards this goal, we designed and implemented multiple visualizations. For instance, the fill level analyzer that can be used to explore station states throughout the day (local perspective) or the network characteristics page that supports the investigation of bike-sharing systems on a global scale.

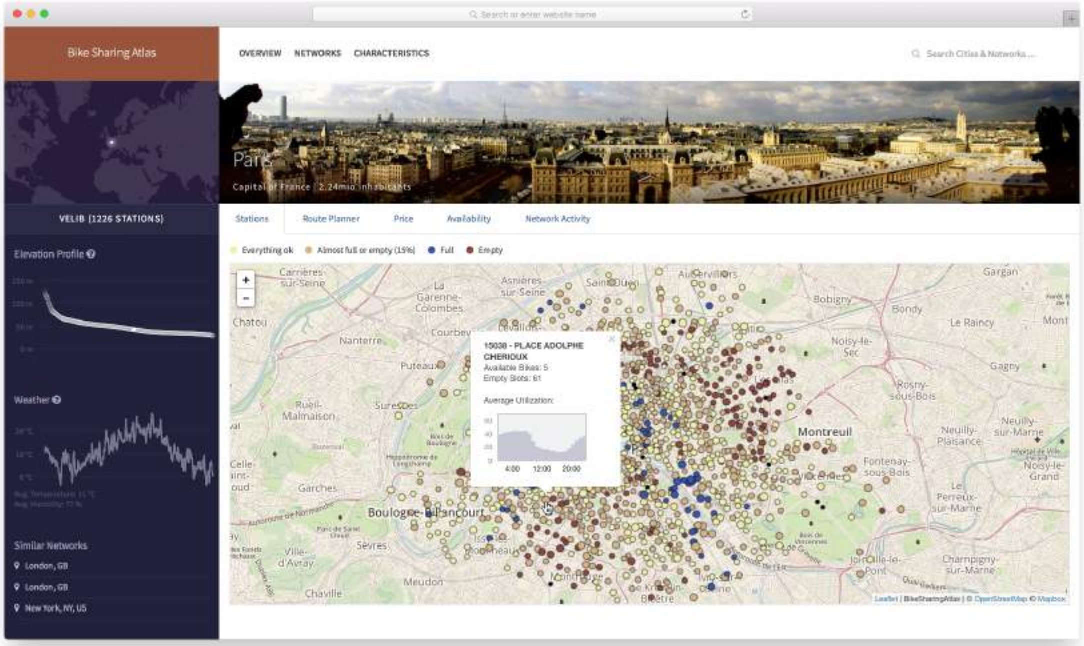


Figure 4.4: Current overview of station fill levels in Paris, France. Tooltips help to predict the availability of bikes and empty docks at certain times during the day.

Fill Level Analyzer

By recording and aggregating station fill levels over a period of 17 months we get an accurate picture of the daily average utilization of each station in a network. The *fill level analyzer*, shown in Figure 4.5, with its multiple coordinated views is based on this data and provides a new approach to identify capacity bottlenecks and commuting behaviors in bike-sharing systems. The multi-series line chart is the core element and shows, for a selected city, the average fill levels (y-axis) during the day (x-axis). Each line represents a station. Due to the different numbers of docking spaces per station, the fill levels are normalized. This contrasts from the work from O'Brien et al. [29] where all station fill levels are aggregated to get a single line. In that case, during the averaging process important information gets lost and it is impossible to expose critical stations that are mostly full or empty. In our proposed system all stations are separated and the user can switch between different modes to explore, for example, fill levels only for weekends or weekdays. Additionally there are two other perspectives: an elevation profile of all stations and an interactive map for providing the geographic context.

Dynamic linking and brushing [38], an interactive visualization technique that connects multiple views, leads to a holistic understanding of the city dynamics. As shown in Figure 4.5-a, the user can draw a line—a hypothetical profile—on top of the multi-series line chart and our algorithm automatically selects similar stations. At the end of the

Figure 4.5: The fill level analyzer view exposes bike-sharing commuting behaviors in London, U.K.

drawn line a slider is displayed and allows the user to specify how many lines she wants to select (default is 20% of the lines). The selected stations are not only highlighted in the line chart (brushing), but also in all other views (linking). This approach allows the user to quickly analyze complex patterns that are distributed over multiple views.

In addition to the multi-series line chart, the lasso tool (freehand selection) can be used within the map to select multiple networks. Thus it is also possible to investigate specific geographical regions (see Figure Figure 5.2-b).

Network Characteristics

To illustrate the global scale of our data, we implemented another multi-coordinated view dashboard that allows users to interactively explore bike-sharing systems globally (see Figure 4.6). Several *frequency charts* (or strip plots), with thin, vertical lines representing individual networks, show the distribution of networks along a set of selected dimensions, such as population, network activity, or the number of reachable stations (see Section 3.4 for more details). Some of the charts contain very similar data points with just a few outliers while others are more equally distributed over the whole range. We have used transparency to make these distributions clearly visible. Areas with a high density of data points show up darker than areas with a low density. The color coding of selected

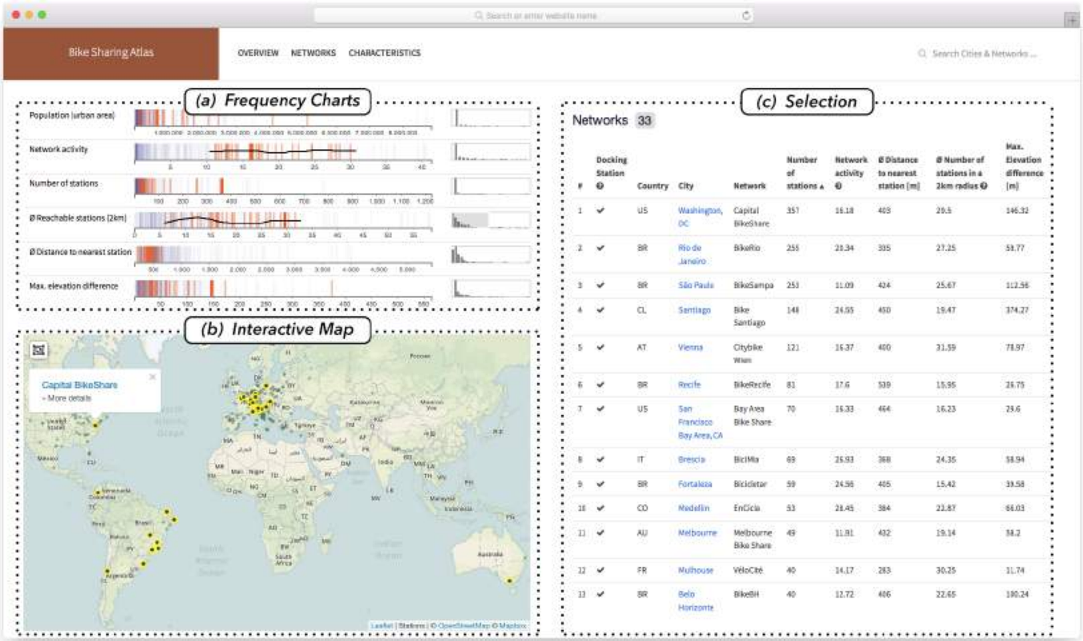


Figure 4.6: Frequency charts linked with a map view allow users to interactively explore and compare network characteristics

networks ensures a strong *pop-out* effect [39]. Similar to the fill level analyzer, users can draw lines on top of these charts to select certain ranges of interest. Again leveraging the linking and brushing approach, the selection is highlighted on all other dimensions, as well as in the corresponding map. A small histogram to the right of each frequency chart serves as an additional indicator of the distribution and can be used to narrow down the range. This concept is called *focus+context* and we have used it multiples times in our interactive visualization system [14]. Although, we have used different chart types and visual encodings it always follows the same idea: detailed information about a selected subset (focus) is embedded in a view that also contains an overview of the entire data structure (context). The information of interest is in the foreground (e.g. frequency chart) while also preserving a global view at reduced detail (e.g. histogram). Vice versa, and in addition to the frequency charts, upon filtering a specific geographical region (lasso selection), the received values are highlighted in the frequency charts and displayed in the table on the right. Particularly when selecting multiple networks at the same time, the sortable table can be used to further analyze these attributes.

The multi-coordinated views, on this page, are based on a small JSON configuration file that specifies their appearance. In case we want to add more network characteristics in the future we can easily append another field to the JSON object and the view will be updated automatically.

The supplemental video illustrates our linking and brushing approach and Section 5.2 discusses usage scenarios associated with the fill level analyzer and network characteristics.

4.3.3 Ease of use

In order to make the tool usable for a broad audience, we cannot have a steep learning curve as is the case for many expert visualization and data analysis tools. Our tool should be self-explanatory and provide easy entry points. We thus sought to consistently use easy to understand visual encodings, specifically as the visualization literacy of the general public is known to be low [40]. While earlier iterations of the tool included more complex encodings, such as a heatmap or cluster views, we eventually settled with an approach that is solely based on maps and simple statistical graphics. Instead of complex visual encodings, we seek to allow for exploratory discoveries through rich, yet easy to understand linking and brushing interactions. The whole user interface is characterized by its minimalist, function-oriented design. The color palette and fonts were chosen with the aim to enhance the readability and the contrasts, and on the basis of common design principles [39]. We also followed the idea of suggested interactivity [36], in the form of little preview videos and tooltips, to guide first-time users through the visualization system.

4.3.4 Modular design

As exploratory data analysis is an inherently ill-defined process, we designed the system to be open and flexible toward future changes. We thus opted for a modular design that can be easily extended with data from new bike-sharing systems or other sources, as well as with different views onto this data. We achieved this flexibility by implementing a client-server architecture that strictly separates the data model from the graphical user interface. Additional data records can be added at any time and will be visible in the frontend without any changes. Most of the D3 visualizations are encapsulated as modules that can be loaded and reused at various places in the web application. The layout of the Bike Sharing Atlas provides only a rough framework and ensures a uniform appearance. New detail views on a single network can be added to the tab navigation and further pages presenting other information can be easily integrated in the primary navigation.

CHAPTER 5

Evaluation

5.1 Expert Interviews

Over the course of the last two years, we evaluated and discussed our approach with the following target groups:

- Bike-sharing operator
- City planners
- Public authorities
- Urban sociology researchers
- General public users

Overall, we interviewed 15 participants in 13 single or group sessions. The interviews were conducted at the University of Vienna or in the office rooms of the respective participants. The procedure varied considerably depending on the interviewees' input and their interests. In general, all the interviews started with a demonstration of the prototype followed by questions and an open discussion. The meetings with the general public users followed a *thinking-aloud* approach [41]. Instead of presenting the tool, the users were asked to perform a few tasks on their own and to talk to me while working on them. This technique allowed me to observe interactions and thought processes of non-expert users. Therefore, think aloud is a commonly used method in the field of human-computer interaction to identify usability issues. Below, we present further details regarding our interviewed target groups.

5.1.1 Bike-Sharing Operator

Since the start of the project in November 2014 the operator of the local bike-sharing network in Vienna accompanied and supported us. The system in Vienna is one of the first of its kind and went through several stages of development. The responsible persons have built up a wealth of experience in the past years. For example, to improve the daily rebalancing operations or to evaluate new locations for docking stations. Although some methods are rather outdated and cumbersome, the company uses various analysis tools and visualizations since the early days. Joint brainstorming sessions, the provision of historical data at the beginning and their reports on the challenges in the setup and operation of a cost-efficient bike-sharing system significantly influenced the evolution of our project. After an initial brainstorming session, we had two further extensive interviews and multiple e-mail exchanges to evaluate our prototypes with these domain experts. In the first follow-up interview, we presented and discussed *CitbikeVis*, an interactive visualization to analyze station bottlenecks and other aspects of a single bike sharing system (for more details, see Chapter 6). In the second one, we proposed the Bike Sharing Atlas and explored together different potentials of our global approach on leveraging bike-sharing data (see Section 5.2).

5.1.2 City Planners

Bike-sharing has changed the urban landscape and commuting behaviors in many cities around the world and thus attracted a lot of interest of city planners and developers. For that reason, we were pleased to hear the opinions of these experts on our project. Despite the dense network of city public transport in Vienna, bike-sharing has been established as a cheap and comfortable alternative. The interview with these experts revealed that comparisons with other cities play an important role. For instance, analyses of other networks revealed that the station density in Vienna is rather low and there are efforts to increase it in the future. In this context, the interviewees were interested in additional features that automatically recommend and analyze new locations and extensions of docking stations. While our system currently works on a macro level with high-level characteristics, an extension or a separate tool could also incorporate the micro level and support these kinds of tasks.

5.1.3 Public Authorities

Local politicians often act as an intermediary between city planners and bike-sharing operators. They are involved in strategic decisions and have to manage the city budget. The interview in May 2016 showed us that they are also highly interested in multi-city comparisons. Furthermore, they appreciate our data-driven and visual approach to understanding city dynamics. This enables them to get an unbiased perspective and to gain various insights without being an expert. Due to the tight budget situation, the bike-sharing expansion in Vienna is currently paused and the interviewees did not see a direct practical benefit at the time of our meeting. In case of an analysis or extension of the current network in the future, they would be very interested in using our web application.

5.1.4 Researchers in Urban Sociology

During the development of the first prototype, we recognized that the Bike Sharing Atlas enables us to identify various interesting behaviors and commuting patterns. After conducting two interviews with sociologists the implications and opportunities of our approach became more evident. Although our system is primarily built on bike-sharing data, the feedback from these researchers illustrated potential use cases for other disciplines as well. Combined with other datasets, such as job density statistics, our system can help to study local effects of residential segregation or to perform cross-city comparisons. It also emerged from the discussion that it is not conducive to overload the web application with many different datasets from external sources. Instead, the integration of an upload functionality for custom datasets (e.g. CSV or GeoJSON files) could further enhance the value of the visualization system.

5.1.5 General Public

In the course of the project, we conducted six *think-aloud* studies with general public users – individuals who do not have an explicit domain knowledge. We asked them to execute several tasks: a) check the live station fill levels of a specific city; b) use the bike-sharing route planner; d) try the fill-level-analyzer and c) explore other networks. Following the think-aloud technique, they should speak aloud all thoughts that came to their mind. The feedback received in the interactions with this target group was very positive and has shown the added value, especially for those traveling or moving to new cities. The user interface and the aesthetics of the whole system were positively emphasized. However, some concerns were raised that the web application is primarily interesting for bike-sharing enthusiasts.

5.2 Usage Scenarios

The insights we gained together with these people informed many design considerations and shaped the final implementation. The usage scenarios below are based on them and are meant to illustrate the potential of our interactive visualization tool.

5.2.1 Usage Scenario: Commuting Patterns

Besides bike-sharing operators, that can use the fill level analyzer, shown in Section 4.3.2, to analyze capacity bottlenecks, such as frequent outages, city planners and public authorities can use the visualization system also to identify and communicate urban commuting behaviors. For example, stations in the city center of London get full during the day and empty through the night, as shown in Figure 4.5. This phenomenon is observable in many cities worldwide as illustrated in Figure 5.1. Milan and New York City show very similar behaviors. People from the outskirts are commuting to downtown areas, such as Fifth Avenue in New York or Piazza Duomo in Milan, in the morning and return in the evening. Mexico City's pattern is a bit more faceted and shows multiple geographical hotspots. Although our two city planner interviewees have hypothesized a considerable separation of residential and commercial spaces in certain cities, our tool exposed visible evidence for this behavior.

The small thumbnails in the middle column of each view show how commuting patterns differ during the week from those on the weekend. Fill levels on Saturdays and Sundays have usually a much flatter profile. In the interactive visualization system, users can click on one of the thumbnails to show it in the main view enlarged.

Barcelona and other Spanish cities show this characteristic morning commute curve too but in the afternoon it becomes vague. We can imagine that this effect might stem from the different working patterns in Spain with a longer lunch break and longer working hours in the evening [42]. This assumption could be further investigated by sociology researchers with the aid of our visualization system.

Other cities, such as Marseille or Vienna, are instead characterized by a mixed-use development, without such a clear commuting pattern in filling levels (i.e., mostly balanced fill levels).

5.2.2 Usage Scenario: Combining with Additional Data

Elevation Profiles

Multiple guided brainstorming sessions with a bike-sharing operator revealed that elevation differences between stations are essential factors for the cost-efficient functioning of a system. Stations at higher altitudes tend to be empty more frequently because users are usually more downhill-oriented. The bikes must be re-balanced manually by the operator. By combining the historical bike-sharing records with elevation profiles, the implemented visualization system also supports a closer investigation of this question. As part of the fill level analyzer, described in the previous usage scenario, we have integrated an *elevation profile* of all stations, which is also connected via linking and brushing to all other views. Instead, of a u-shaped commuting pattern such as in Figure 5.1-a or 5.1-b, average filling levels in Vienna, for example, remain mostly constant throughout the day but show clear evidence of fewer available bikes at higher altitudes. The screenshot in Figure 5.2 shows this scenario visually. Stations are highlighted in the multi-series line-chart, in the map, and shown as filled white circles in the elevation profile on the left side. Despite the relatively small elevation difference of 78m between the highest and lowest station its impact must be considered when planning new stations in Vienna.

The maximum elevation difference between the highest and lowest station serves as one of our network characteristics and can be also used to analyze elevation patterns on a global scale.

Weather Records

Besides the elevation profiles, we further enhanced our database with hourly weather records for all cities. In previous works weather has been found a substantial factor in

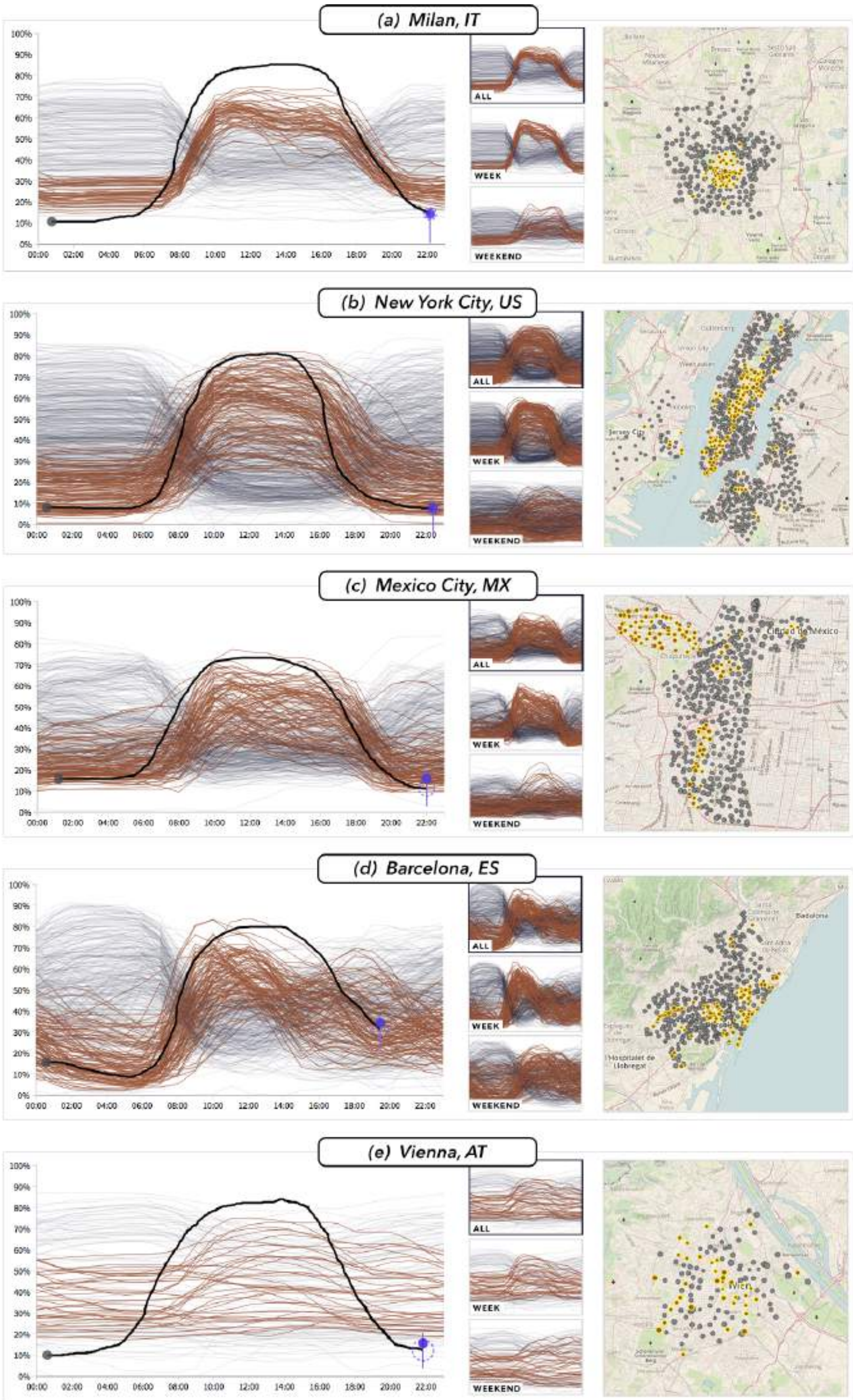
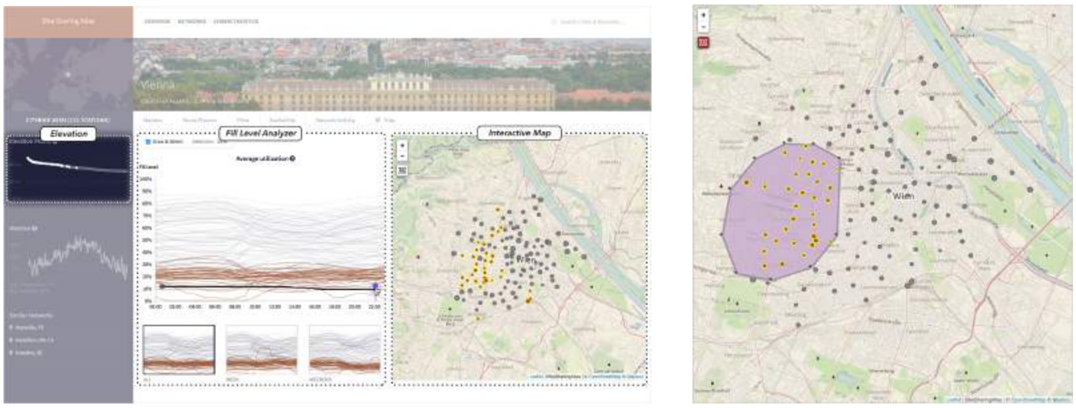


Figure 5.1: Station fill levels visualized as multi-series line charts serve as indicators of commuting behaviors worldwide. Small thumbnails in the middle column of each view show how patterns differ during the week from those on the weekend. Users can click on one of the thumbnails to show it in the main view enlarged. Stations can be selected either in the line chart or in the map view and are highlighted respectively.



(a) Fill level analyzer shows mostly balanced fill levels in Vienna but provides evidence of fewer available bikes at higher altitudes

(b) Lasso selection tool can be used to filter stations based on a geographical region

Figure 5.2: Analyzing station fill levels of Vienna's bike-sharing network

bike sharing demand [43, 44]. In addition to cross-correlation between the temperature and the network activity, which we compute for easy global comparisons, we also added an additional *time series view* showing this data visually. Vienna, for example, has a distinctive pattern and a particularly strong relationship between the network activity and the temperature, as shown in Figure 5.3. The two white cuts in April and August were caused by server issues and do not represent the actual network activity. Visualization is also good in quickly revealing such anomalies [45]. Analyzing the impact of weather conditions on bike-sharing can help to improve rebalancing operations and the planning of new systems in the future [44].

Demographic Information

Generally, we opted for a modular design that can be easily extended with data from other sources, as well as with different views onto this data. Especially future work in urban sociology would benefit from an integration of additional context information, such as demographic developments (workplace density, gross domestic product etc).

For instance, we integrated a choropleth map showing Vienna's population density [1] in an early stage prototype, as illustrated in Figure 5.4. Choropleth maps display geographical regions, such as subdistricts in our case, that are colored in relation to the underlying data values. This visualization type is very popular to present quantitative values across geographical regions but if not used correctly it can be misleading and distort the interpretation. Instead of visualizing absolute values we have used ratios (or derived values) to also consider the varying sizes of the areas. However, currently it is

[1] Data source:
https://github.com/anitagraser/Webmapping-Sandbox/tree/gh-pages/data

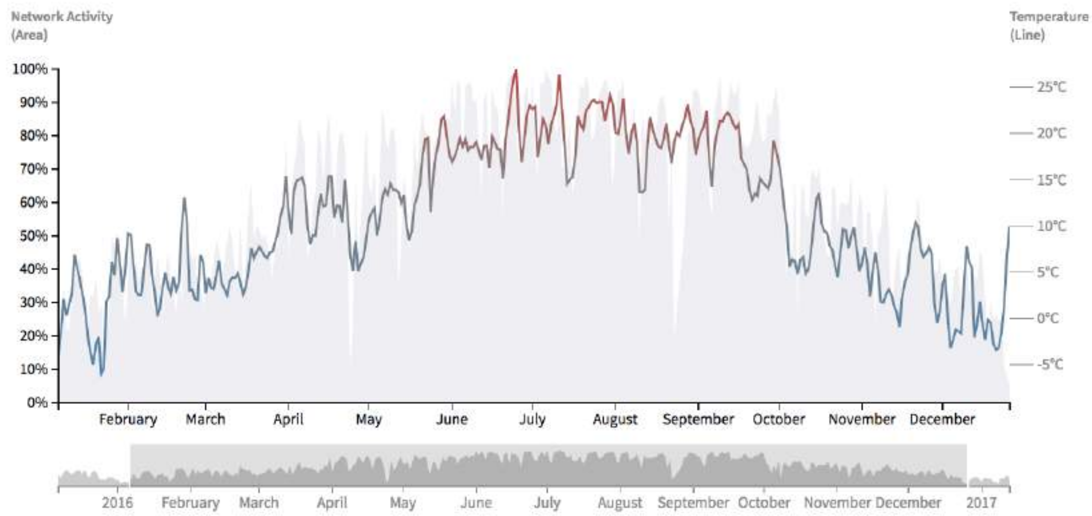


Figure 5.3: Time series showing the network activity (gray area) and the temperature (line) in Vienna

very difficult to get demographic data on this level of detail from many cities worldwide. Therefore, this feature is not included in the current version of *bikesharingatlas.org* but in the future our system could be extended to provide an upload option for custom data sets.

5.2.3 Usage Scenario: Multi-City Comparisons

In order to provide an overview, a shared global perspective on the collected data, we implemented multiple interactive visualizations of all networks, such as zoomable maps or a small multiples view (see Figure 4.2-a and 4.2-b). Due to the same scales and axes, *small multiples* are very efficient for giving a first quick overview and for comparison purposes [46]. The individual vector maps with bike-sharing stations as dots serve as indicators of the size, structure, and density of networks. Users can browse through these networks and get further details on demand.

As described in Section 4.3.2, we implemented another page with multi-coordinated views to explore network characteristics and to further leverage the global scale of the data. Bike-sharing networks can be filtered and analyzed along multiple dimensions by using the frequency charts. The lasso tool within the interactive map allows users to focus the analysis on a certain geographic region, such as North America.

While operators of these systems have mostly a rather narrow and local view, our interactive visualization system enables them to explore, compare, and learn from other networks worldwide. Similarly, politicians and transportation authorities who are planning

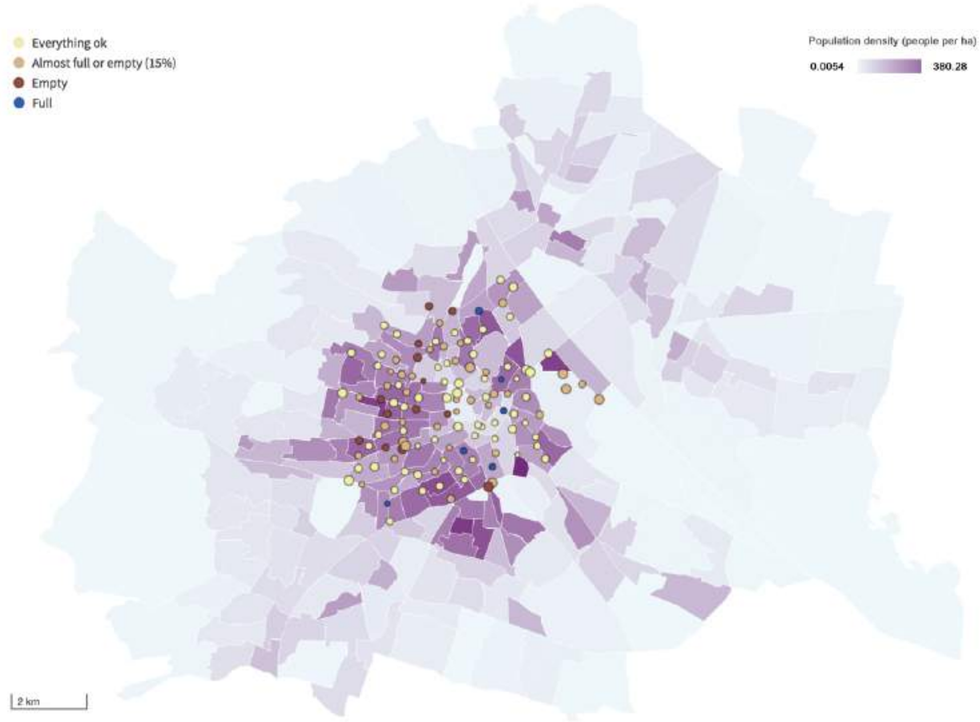


Figure 5.4: Choropleth map of the population density in Vienna. The colored circles show live station fill levels on a weekday at 2pm.

a new bike-sharing system can get interesting insights from equal-sized cities, such as the number of reachable stations in a 2 kilometer radius.

5.2.4 Usage Scenario: General Public

By bringing together distributed, heterogeneous data on a single platform we also simplify the access for general public users. While this user group is often not interested in in-depth analysis, it can benefit from a lightweight interface that leverages this rich data source. Our implemented system includes various features that illustrate how typical tasks that appeal to the general public can be supported, for example, maps showing live station fill levels or a bike-sharing route planner for all our 468 networks, shown in Figure 4.4 and Figure 5.5 respectively.

Within the *route planner* users can enter trip start- and endpoints and the system automatically finds the nearest stations and the fastest route. More precisely, we calculate two walking routes and one cycling route for each request that are combined

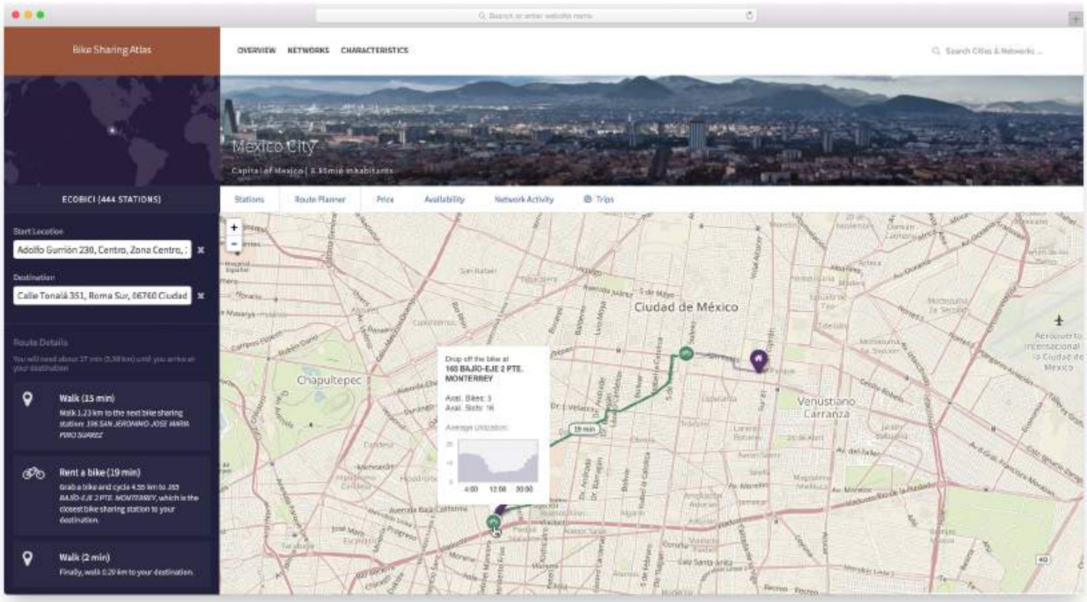


Figure 5.5: The route planner shows the fastest bike-sharing route between two endpoints in Mexico City

together into one itinerary [2]. First, we search for the fastest walking route between the starting point and the nearest bike-sharing station. Second, we compute cycling directions from this point to the closest station at the desired destination and third, we calculate the walking route for the last stretch to the final destination. The result is shown visually in the map view and in textual form in the left sidebar. Additional tooltips in the interactive map show the current fill level and a historical profile for each station, which we computed by averaging across the 17 months of data that we recorded. This information can be used to roughly predict how the availability might look like at a certain point in time, similar to Google's 'popular times' feature [47]. While similar planning tools exist for many systems [48, 49], our data allows a unified approach across them.

5.3 Quantitative Analysis of User Behavior

Soon after finishing our first prototype we published our web-based visualization system under the domain name *http://bikesharingatlas.org* with two main objectives: first, to give our expert interviewees the chance to try the system extensively without our guidance and second, to make it accessible to as many people as possible.

[2] Walking and cycling routes are computed by using the *Mapbox Directions* service (https://www.mapbox.com/directions/)

In this context, we were especially interested in the following questions:

1. Which pages and features create more interest than others?
2. How long people usually spend there and what is the average bounce rate?
3. How do users interact with the system as a whole and with the introduced interaction techniques in detail?
4. Are there any bugs or barriers that affect the user experience?
5. Where do people come from geographically and what language do they speak?

To answer these questions, we have integrated an analysis framework that contains the tracking tools *Google Analytics* [3] and *Mouseflow* [4], and an online feedback form that is displayed to users after two minutes of staying on the page (similar to Figure 5.7). Google Analytics is one of the most widely used tracking tools for web projects. It can be easily integrated into any website and it provides us with detailed information, such as traffic sources, page popularity, which devices and browsers are used, and many other statistics that help us to get more insights into the user behaviors on our platform. Mouseflow is different in the way it processes and presents the tracking data. In addition to classical statistical methods, Mouseflow creates recordings of each visit and we can watch the users' interactions as a video (screencast) afterwards. Furthermore, it generates various *heatmaps* of all visited subpages and shows aggregated user interactions visually. As shown in Figure 5.6, the tool creates an overlay on top of the web application that displays, for example, the number of clicks on each element. This feature can help us to evaluate which filters are getting applied, which networks are especially interesting or how often do people click on certain menu items.

In the period September 2016 to March 2017, we recorded 741 different users and 5,449 individual page views in total. The average length of a session was roughly 5 minutes and users visited 3.97 pages during one session on average. Most users came from Austria (27.4%), followed by the United States (20.51%), Bosnia Herzegovina (8.64%), and the United Kingdom (7.96%). The majority of people used Google Chrome (73.01%) and a desktop computer (86.37%) to browse through the visualization system.

Unfortunately, the survey response rate for the feedback questions—shown to users within a popup on the website—was very small. In general, the total number of users was also rather small for this long time period. Thus, we put the primary focus of our evaluation on the interviews.

Media coverage on *futurezone* [1] and in the *uni:view magazine* [2] validates the general interest in the Bike Sharing Atlas.

[3] https://analytics.google.com
[4] https://mouseflow.com

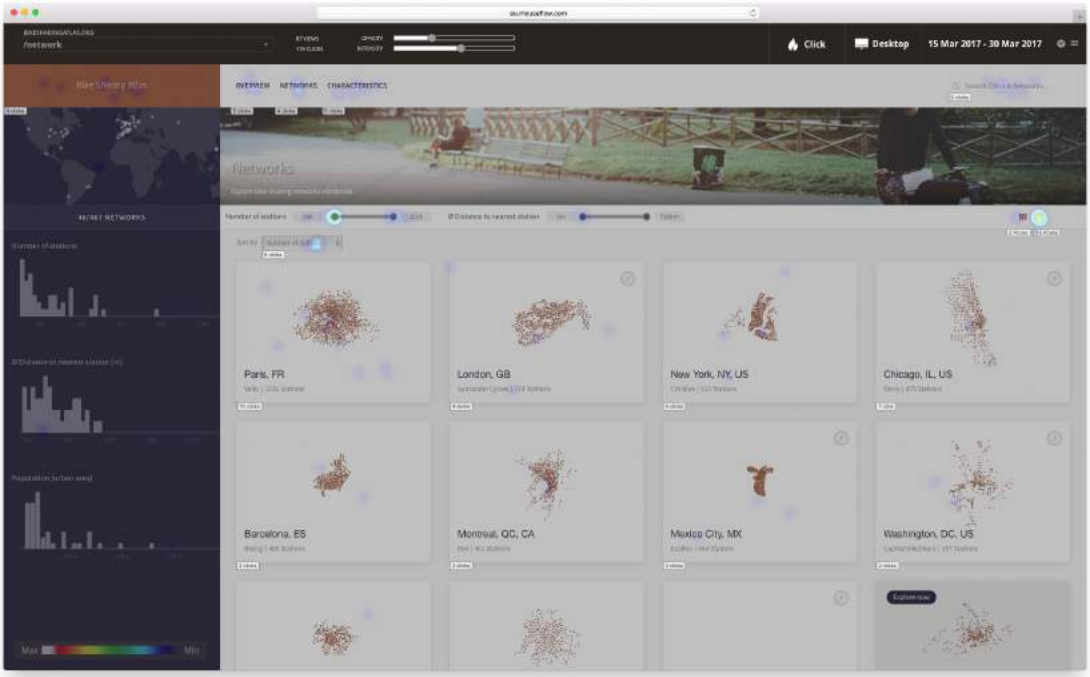


Figure 5.6: Heatmaps help to analyze user behaviors by overlaying click interactions

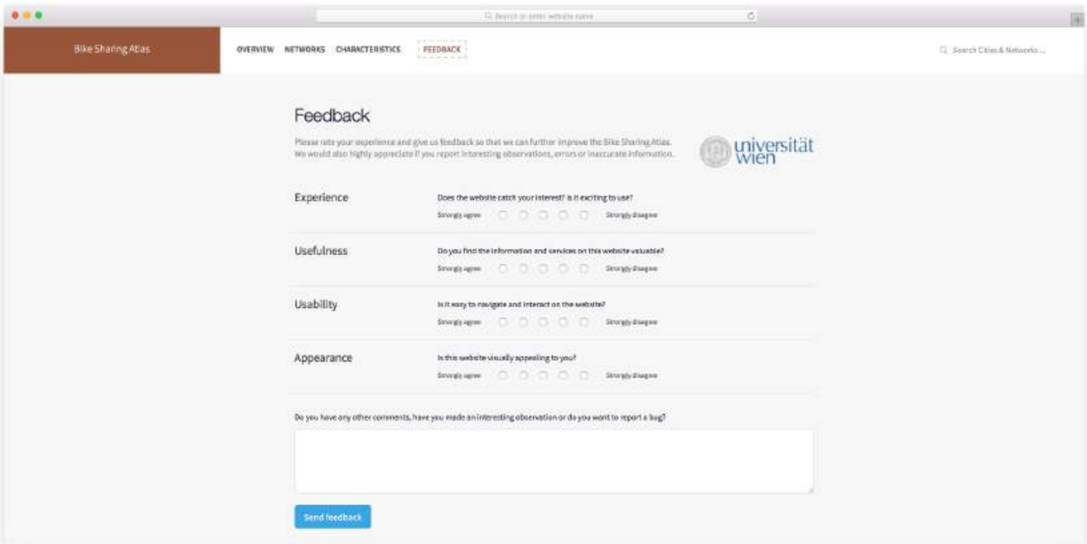


Figure 5.7: Screenshot of the feedback form that was integrated to collect opinions from general public users

CHAPTER 6

Discussion

6.1 Foundations & Process

The current methodological approach in visualization research often follows a design study. Sedlmair et al. [50] defined it “as a project in which visualization researchers analyze a specific real-world problem faced by domain experts, design a visualization system that supports solving this problem, validate the design, and reflect about lessons learned in order to refine visualization design guidelines”.

From this point of view, our approach is different, unconventional and more related to the parallel multi-channel approach to visualization from Wood et al. [5]. Although, our project was initiated in coordination with a bike-sharing operator (domain expert) we did no face a particular problem. Instead, we were confronted with a versatile dataset—produced by Vienna’s bike-sharing system over several years—and reports on the challenges of setting up and operating a bike-sharing network.

As shown in the project timeline in Figure 6.1 we first implemented CitybikeVis, an interactive visualization system for the exploration and analysis of Vienna’s bike-sharing system. Multi-coordinated views provided different perspectives on the data, such as station fill levels (see Figure 6.2), trips (see Figure 6.3) or damage reports. After another interview with the bike-sharing operator and a demonstration of our tool, it became more and more clear that this type of visualization can support and accelerate various analysis processes but it will not reveal entirely new insights. The operator has a wealth of experience and knows popular routes, station bottlenecks and other dynamics within the local network very well. On the other hand, we could observe that the operator, and published visualization tools for other cities take a rather narrow view that focuses on a local network. These findings led to our idea of a global, data-driven approach to understanding smart cities.

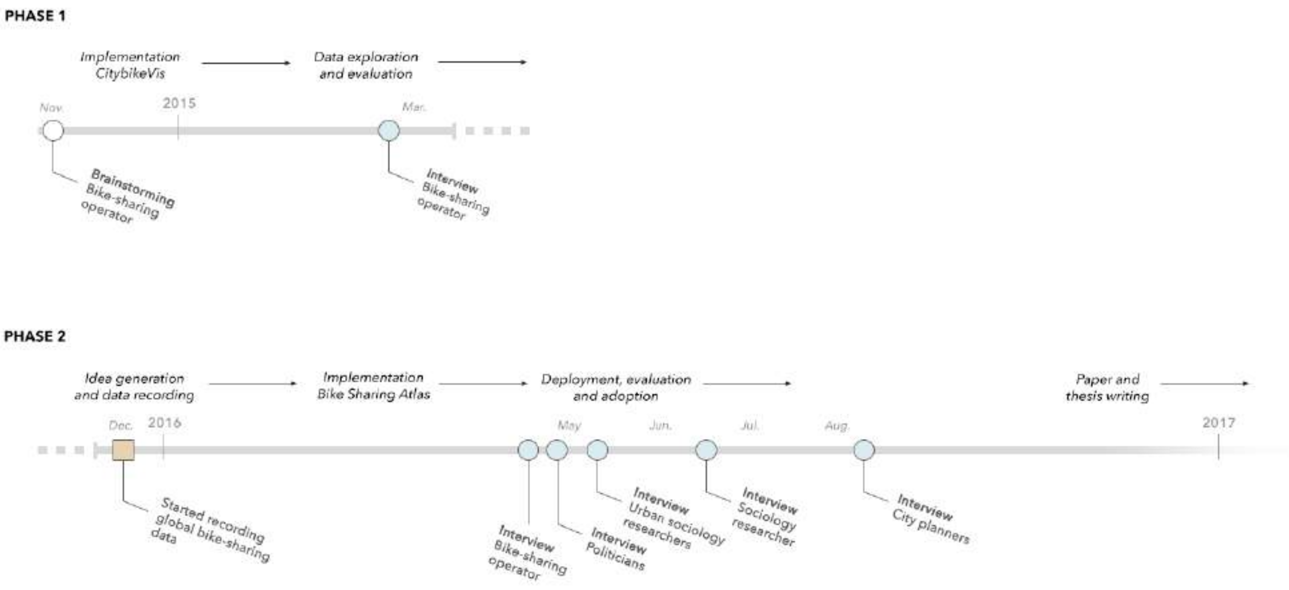


Figure 6.1: Project timeline.

In the second phase of our project (Figure 6.1), beginning in December 2015, we started recording live station fill levels of several hundred bike-sharing networks worldwide. Within a short period of time, we accumulated several million data records that provided the basis for the implementation of our visualization system, *bikesharingatlas.org*.

At the beginning, it was largely unclear how these aspects break down into novel and actionable solutions in terms of data analysis and visualization tools. During an iterative design process, several formative and summative assessments were conducted as interviews and think-aloud studies. Our understanding about the potential and challenges of our approach evolved and the feedback from the evaluations strongly shaped the final implementation.

As described in the usage scenarios above, multiple benefits of a global view on smart city data and urbanization emerged for different target groups. This flexibility and the design of our *general public system* comes also with drawbacks. We deliberately excluded more complex concepts and encodings, such as heatmaps or cluster views, that could be of great interest to expert users. In order to provide these expert features, we would need an advanced mode or a separate user interface.

Figure 6.2: *CitybikeVis*: Analyzing station fill levels of Vienna's bike-sharing system. (a) The control panel allows users to filter data, to change the visual encoding and to select an aggregation mode. (b) The heatmap shows all stations and their aggregated *fill levels* over a certain period of time. Each line represents one station and each column displays the station state, either for a day, week, or month. The cells are color-coded based on the selected variable. In this screenshot we get an high-level overview of all stations in Vienna on a day-level. The matrix is sortable in both dimensions which enables users to identify station bottlenecks and other unusual behavior. (c) The sidebar shows another view on the station states or more details if the user selects a specific cell.

We have also implemented a prototype of such an expert tool that allows users more detailed multi-city comparisons. Figure 6.4 and 6.5 show screenshots of the window-based visualization system that is built on the Bike Sharing Atlas. Instead of several sub-pages that facilitate diverse perspectives on the bike-sharing data, the core of our expert tool is a flexible drawing tool. Users can select a certain city and available views are displayed accordingly in the left sidebar. These views can be positioned freely on the drawing area and thus enable users to compose individual dashboards with multiple visualizations. Identical station IDs in different views are automatically detected and visually emphasized if a user hovers over an element. Currently, it is an early prototype but in the future, it can be further extended to serve as a framework for visualizing data from other smart city sensors.

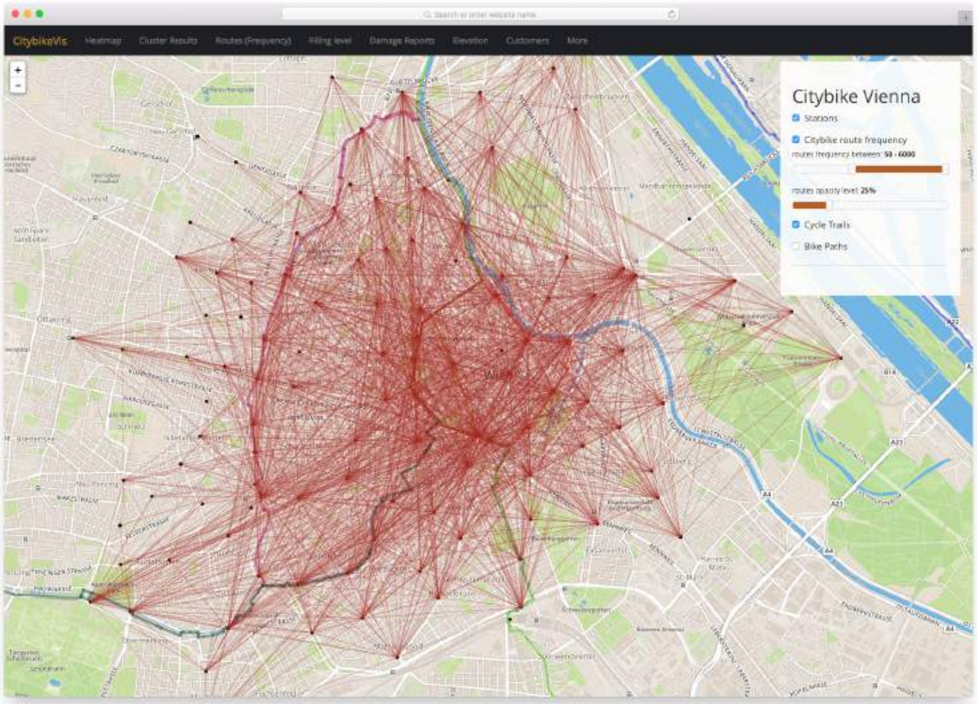


Figure 6.3: Screenshot of the tool *CitybikeVis* focusing on completed trips within Vienna's bike-sharing network. Users can filter the data, show additional overlays, such as cycling paths, and explore popular routes and stations interactively.

6.2 Lessons Learned

In the course of our two-year long project, we ran through multiple development iterations and evaluations with potential target groups. In this section, we briefly reflect on valuable lessons learned from collaborations and from implementing our platform.

6.2.1 Open Data Access

Access restrictions still represent the greatest barrier for leveraging smart city data. While there emerged numerous open data initiatives and online libraries in recent years we realized multiple times how difficult it is to get global data that is similarly structured. Datasets, such as demographic statistics on a city level, are often provided only for specific countries or larger cities. Bike-sharing networks are generally publishing at least their station fill levels but especially operators in Asia restrict access to this information at the moment. Unfortunately, due to the rapidly growing number of bike-sharing networks in Asia, this influenced also the results of this thesis. Similarly, weather services around the world provide easy access to live weather records but the utilization of historical data is restricted behind paywalls. These access limitations are often justified with the argument to protect citizens' rights to privacy. For example, if we have detailed information about completed bike-sharing trips in a small town we would be able to track specific people and reconstruct their movements. Although this personal information must be protected unconditionally, we observed in the course of our project that the public access to global data is often hampered by national barriers or inconsistent data formats. This thesis and other data-driven smart city projects demonstrate the potential and the benefits of making this data accessible and understandable for various target groups and we hope that it will contribute to the growing open data movement.

6.2.2 Custom Visualizations

In the course of the development of multiple prototypes, we had to decide which technologies are most appropriate and effective for the given requirements. We quickly agreed to build an online visualization system built on web technologies. However, the question raised if we can implement the whole system within the *Tableau* [1] environment or if we should use *D3* [2] or other visualization libraries. Tableau is a visualization software that helps users to explore and analyze relational data sets. Interactive dashboards with multiple views and filters can be plugged together and integrated into any website. We have used theses features in various project stages to create and evaluate medium-fidelity prototypes. The assessment of advantages and disadvantages in relation to the JavaScript library D3 quickly revealed that we would lose a lot of flexibility if we solely used Tableau. It is limited to the provided chart templates and in particular, the implementation of custom interaction techniques (e.g. drawing mechanism for the fill level analyzer) would not be possible. The Bike Sharing Atlas was therefore exclusively built with D3 and on the basis of JavaScript.

6.2.3 Domain Experts

Another lesson we have learned is the importance of domain experts for these types of projects. With their specific domain knowledge and their feedback on the prototypes, they are a decisive factor towards the success of a visualization project. In this context, engagement is another crucial factor. First, we cannot expect domain experts delivering us with well-defined problems that are simultaneously interesting visualization research questions, and second—as Sedlmair et al. [50] mentioned as one common pitfall for design studies—there is often not enough time available for activities such as problem analysis, design discussions or evaluations. Although all of our interviewees saw the potential of our system, most of them did not see an immediate need for their daily work or their current projects and, thus, did not accompany our project with a significant time commitment. Instead, they expected from us to further develop the system and to provide them with valuable insights.

[1] https://tableau.com
[2] https://d3js.org

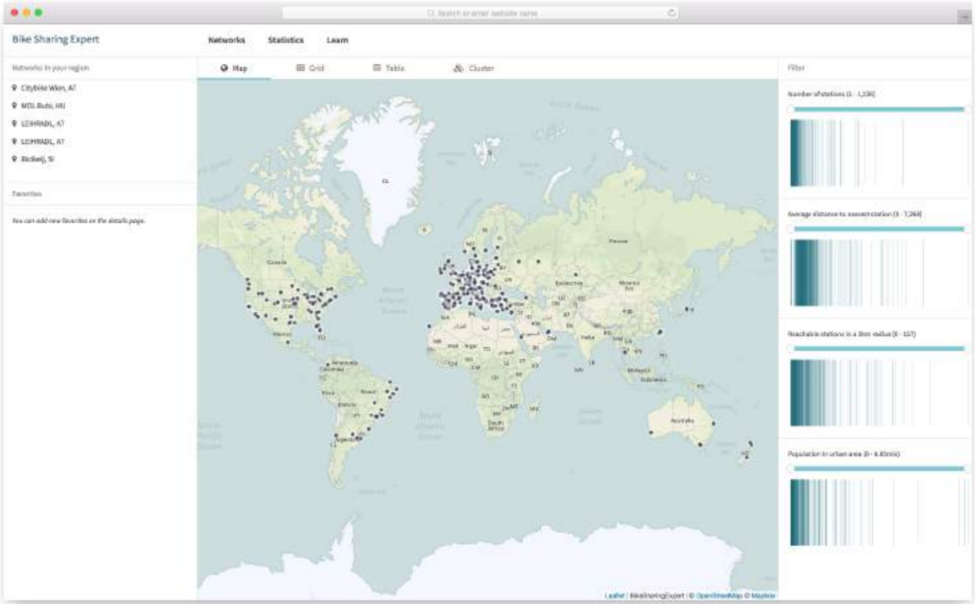


Figure 6.4: Screenshot of the tool *Bike Sharing Expert* showing an overview of all available bike-sharing networks. Users can switch between different perspectives and the frequency charts in the right sidebar can be used to narrow down the search for interesting cities.

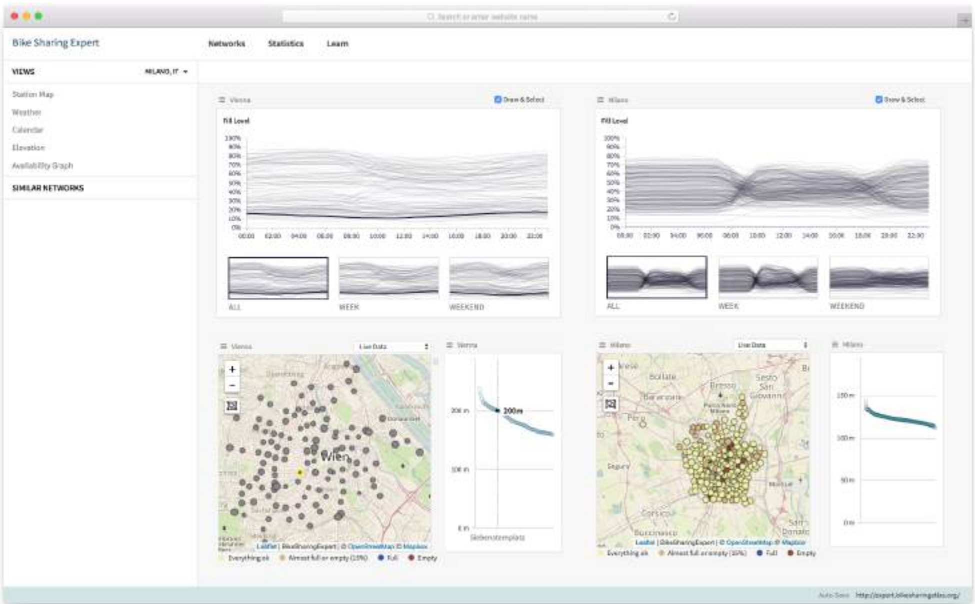


Figure 6.5: Comparison of the bike-sharing networks in Milan and Vienna. Users can select a city in the left sidebar and available views are automatically displayed. These views (e.g. elevation profile) can be added at any position on the drawing area. This flexible, window-based interface allows users to quickly build dashboards and to compare multiple cities.

CHAPTER 7

Conclusions and Future Work

In this work we discuss a data-driven and visual approach to understanding and leveraging smart city data. Through our iterative design process, we found evidence that such an approach can benefit different target groups. The fill level analyzer with its multi-coordinated views, for instance, provides a new way to explore and communicate commuting behaviors in 45 countries. The combination with other data sources can help, for instance, urban sociology researchers to analyze effects of residential segregation. Elevation profiles support bike-sharing operators in identifying bottlenecks with stations at higher altitudes. An integrated route planner and live station fill levels offer a benefit for general public users.

While our system is primarily built around global bike-sharing data we believe that the proposed visual approach is relevant for other smart city sensors too. Observed more closely, we use the data not only to analyze cycling behavior or to build a route planner but also as a way to understand high-level city dynamics more generally. The concept of using sensors for monitoring tasks for which they were not initially designed is called *opportunistic sensing* (or citizen sensing) [51]. Massive amounts of data, produced by car sharing services, taxis, public transport systems, smart meters, or other sensors provide abstractly similar challenges and opportunities. By recording them over a long time period and by making them accessible and visually explorable it could open up new possibilities in understanding and improving urban environments.

So far, we have only scratched the surface and there are many more usage scenarios that could be explored in this context. But also the ones that we have identified would benefit from being complemented, for instance, by further design studies and in-depth collaboration with specific user groups [50]. Our work of gathering and making the data available now provide the first steps towards such future endeavors. Beyond that, we also hope that our work will inspire researchers and designers of other urban data solutions.

List of Figures

List of Tables

Bibliography

[1] D. Kotrba, “Bike Sharing Atlas zeigt, wo die Mieträder stehen,” April 29, 2017. [Online]. Available: https://futurezone.at/science/bike-sharing-atlas-zeigt-wo-die-mietraeder-stehen/260.959.978

[2] T. Dirtl, “Mit dem Radl von Stadt zu Stadt,” April 18, 2017. [Online]. Available: http://medienportal.univie.ac.at/uniview/forschung/detailansicht/artikel/mit-dem-radl-von-stadt-zu-stadt/

[3] P. DeMaio and J. Gifford, “Will smart bikes succeed as public transportation in the United States?” *Journal of Public Transportation*, vol. 7, no. 2, pp. 1–15, 2004.

[4] J. Froehlich, J. Neumann, and N. Oliver, “Sensing and predicting the pulse of the city through shared bicycling,” in *IJCAI*, vol. 9, 2009, pp. 1420–1426.

[5] J. Wood, R. Beecham, and J. Dykes, “Moving beyond sequential design: Reflections on a rich multi-channel approach to data visualization,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 2171–2180, 2014.

[6] J. C. García-Palomares, J. Gutiérrez, and M. Latorre, “Optimizing the location of stations in bike-sharing programs: a GIS approach,” *Applied Geography*, vol. 35, no. 1, pp. 235–246, 2012.

[7] L. M. Martinez, L. Caetano, T. Eiró, and F. Cruz, “An optimisation algorithm to establish the location of stations of a mixed fleet biking system: An application to the city of Lisbon,” *Procedia - Social and Behavioral Sciences*, vol. 54, pp. 513–524, 2012.

[8] T. Raviv, M. Tzur, and I. A. Forma, “Static repositioning in a bike-sharing system: models and solution approaches,” *EURO Journal on Transportation and Logistics*, vol. 2, no. 3, pp. 187–229, 2013. [Online]. Available: http://dx.doi.org/10.1007/s13676-012-0017-6

[9] M. Rainer-Harbach, P. Papazek, B. Hu, and G. R. Raidl, “Balancing bicycle sharing systems: A variable neighborhood search approach,” in *European Conference on Evolutionary Computation in Combinatorial Optimization (EvoCOP)*. Springer, 2013, pp. 121–132.

[10] C. Kloimüllner, P. Papazek, B. Hu, and G. R. Raidl, "Balancing bicycle sharing systems: an approach for the dynamic case," in *European Conference on Evolutionary Computation in Combinatorial Optimization*, 2014, pp. 73–84.

[11] J.-R. Lin and T.-H. Yang, "Strategic design of public bicycle sharing systems with service level constraints," *Transportation research part E: logistics and transportation review*, vol. 47, no. 2, pp. 284–294, 2011.

[12] M. C. Guenther and J. T. Bradley, "Journey data based arrival forecasting for bicycle hire schemes," in *International Conference on Analytical and Stochastic Modeling Techniques and Applications.* Springer, 2013, pp. 214–231.

[13] P. Borgnat, P. Abry, P. Flandrin, C. Robardet, J.-B. Rouquier, and E. Fleury, "Shared bicycles in a city: A signal processing and data analysis perspective," *Advances in Complex Systems*, vol. 14, no. 03, pp. 415–438, 2011.

[14] T. Munzner, *Visualization Analysis and Design.* CRC Press, 2014.

[15] G. Andrienko, N. Andrienko, P. Bak, D. Keim, and S. Wrobel, *Visual analytics of movement.* Springer Science & Business Media, 2013.

[16] R. Beecham, J. Wood, and A. Bowerman, "Studying commuting behaviours using collaborative visual analytics," *Computers, Environment and Urban Systems*, vol. 47, pp. 5–15, September 2014.

[17] G. N. Oliveira, J. L. Sotomayor, R. P. Torchelsen, C. T. Silva, and J. L. Comba, "Visual analysis of bike-sharing systems," *Computers & Graphics*, vol. 60, pp. 119–129, 2016.

[18] J. Corcoran, T. Li, D. Rohde, E. Charles-Edwards, and D. Mateo-Babiano, "Spatio-temporal patterns of a public bicycle sharing program: the effect of weather and calendar events," *Journal of Transport Geography*, vol. 41, pp. 292–305, 2014.

[19] B. Wellington, "Mapping Citi Bike's riders, not just rides," 2014, last accessed April 26, 2017. [Online]. Available: http://iquantny.tumblr.com/post/81465368612/mapping-citi-bikes-riders-not-just-rides

[20] H. R. Alberts, "How New Yorkers and tourists use Citi Bike on two nice days," 2014, last accessed April 26, 2017. [Online]. Available: http://ny.curbed.com/2014/3/31/10122748/how-new-yorkers-and-tourists-use-citi-bike-on-two-nice-days

[21] A. Woodruff, "Bostonography: Hubway trip explorer," 2012, last accessed April 26, 2017. [Online]. Available: http://bostonography.com/hubwaymap/

[22] V. Chiraphadhanakul, "Bay area bike share: Data challenge," 2014, last accessed April 26, 2017. [Online]. Available: http://babs.virot.me/

[23] S. Jacobsen, “Who’s faster? Divvy riders vs transit riders,” 2015, last accessed April 26, 2017. [Online]. Available: http://transitized.com/divvy-vs-transit/

[24] M. Z. Austwick, O. O’Brien, E. Strano, and M. Viana, “The structure of spatial networks and communities in bicycle sharing systems,” *PLoS ONE*, vol. 8, no. 9, p. e74685, 2013.

[25] A. Bargar, A. Gupta, S. Gupta, and D. Ma, “Interactive visual analytics for multi-city bikeshare data analysis,” in *3rd International SIGKDD Workshop on Urban Computing (UrbComp)*, 2014, short paper no. 5.

[26] T. Nagel, C. Pietsch, and M. Dörk, “Staged analysis: From evocative to comparative visualizations of urban mobility,” in *Proceedings of the IEEE VIS Arts Program, VISAP*, 2016, pp. 23–30.

[27] O. O’Brien, “Bike share map,” September 2017, last accessed April 26, 2017. [Online]. Available: http://bikes.oobrien.com

[28] R. Meddin and P. DeMaio, “The bike-sharing world map,” September 2017, last accessed April 26, 2017. [Online]. Available: http://bikesharingmap.com

[29] O. O’Brien, J. Cheshire, and M. Batty, “Mining bicycle sharing data for generating insights into sustainable transport systems,” *Journal of Transport Geography*, vol. 34, no. 262–273, 2014.

[30] T. von Landesberger, F. Brodkorb, P. Roskosch, N. Andrienko, G. Andrienko, and A. Kerren, “Mobilitygraphs: Visual analysis of mass mobility dynamics via spatio-temporal graphs and clustering,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 11–20, 2016.

[31] L. Yu, W. Wu, L. Xiaohui, L. Guangxia, W. Siong Ng, S.-K. Ng, Z. Huang, A. Arunan, and H. M. Watt, “iVizTRANS: Interactive visual learning for home and work place detection from massive public transportation data,” in *IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2015, pp. 49–56.

[32] N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva, “Visual exploration of big spatio-temporal urban data: A study of New York City taxi trips,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2149–2158, 2013.

[33] F. Miranda, H. Doraiswamy, M. Lage, K. Zhao, B. Gonçalves, L. Wilson, M. Hsieh, and C. T. Silva, “Urban pulse: Capturing the rhythm of cities,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 791–800, 2017.

[34] CityBikes API, 2017, last accessed April 26, 2017. [Online]. Available: https://api.citybik.es

[35] C. M. de Chardon and G. Caruso, "Estimating bike-share trips using station level data," *Transportation Research Part B: Methodological*, vol. 78, pp. 260–279, 2015.

[36] J. Boy, L. Eveillard, F. Detienne, and J.-D. Fekete, "Suggested interactivity: Seeking perceived affordances for information visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 639–648, 2016.

[37] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in *IEEE Symposium on Visual Languages*, 1996, pp. 336–343.

[38] J. Heer and B. Shneiderman, "Interactive dynamics for visual analysis," *Queue*, vol. 10, no. 2, pp. 1–26, 2012.

[39] C. Ware, *Visual thinking For design.* Morgan Kaufmann, 2008.

[40] J. Boy, R. A. Rensink, E. Bertini, and J.-D. Fekete, "A principled way of assessing visualization literacy," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 1963–1972, 2014.

[41] C. Lewis and J. Rieman, *Task-centered User Interface Design: A Practical Introduction.* University of Colorado, Boulder, Department of Computer Science, 1993.

[42] S. Jones, "Working 9 to 8: Spain seeks to shorten 11-hour working day," December 2016, last accessed April 26, 2017. [Online]. Available: https://www.theguardian.com/world/2016/dec/13/spain-leaves-franco-in-past-as-it-seeks-to-move-clocks-back-an-hour

[43] M. S. Smith and G. Kauermann, "Bicycle commuting in Melbourne during the 2000s energy crisis: A semiparametric analysis of intraday volumes," *Transportation Research Part B: Methodological*, vol. 45, no. 10, pp. 1846 – 1862, 2011.

[44] K. Gebhart and R. B. Noland, "The impact of weather conditions on bikeshare trips in Washington, DC," *Transportation*, vol. 41, no. 6, pp. 1205–1225, 2014.

[45] S. Kandel, A. Paepcke, J. Hellerstein, and J. Heer, "Wrangler: Interactive visual specification of data transformation scripts," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 2011, pp. 3363–3372.

[46] E. Tufte, *Envisioning Information.* Cheshire, CT, USA: Graphics Press, 1990.

[47] Google, "Popular times," 2017, last accessed April 26, 2017. [Online]. Available: https://goo.gl/F7D0c9

[48] Citymapper, "The ultimate transport app," 2017, last accessed April 26, 2017. [Online]. Available: https://content.citymapper.com

[49] Mapbox, "Plan a ride with surface, directions, and turf.js," 2017, last accessed April 26, 2017. [Online]. Available: https://www.mapbox.com/bites/00094/

[50] M. Sedlmair, M. Meyer, and T. Munzner, “Design study methodology: Reflections from the trenches and the stacks,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2431–2440, 2012.

[51] D. Havlik, S. Schade, Z. A. Sabeur, P. Mazzetti, K. Watson, A. J. Berre, and J. L. Mon, “From sensor to observation web with environmental enablers in the future internet,” *Sensors*, vol. 11, no. 4, pp. 3874–3907, 2011.