IDENTIFYING USAGE PROFILES OF STATION-BASED CAR-SHARING MEMBERS USING CLUSTER ANALYSES

Anna Reiffer (corresponding author)
Institute for Transport Studies, Karlsruhe Institute of Technology (KIT)
Kaiserstrasse 12, 76131 Karlsruhe, Germany
Tel: +49 721 6084 7735
Email: anna.reiffer@kit.edu

Tim Wörle
Institute for Transport Studies, Karlsruhe Institute of Technology (KIT)
Kaiserstrasse 12, 76131 Karlsruhe, Germany
Tel: +49 721 6084 7733
Email: tim.woerle@kit.edu

Lars Briem
Institute for Transport Studies, Karlsruhe Institute of Technology (KIT)
Kaiserstrasse 12, 76131 Karlsruhe, Germany
Tel: +49 721 6084 7772
Email: lars.briem@kit.edu

Tamer Soylu
Institute for Transport Studies, Karlsruhe Institute of Technology (KIT)
Kaiserstrasse 12, 76131 Karlsruhe, Germany
Tel: +49 721 6084 3048
Email: tamer.soylu@kit.edu

Martin Kagerbauer
Institute for Transport Studies, Karlsruhe Institute of Technology (KIT)
Kaiserstrasse 12, 76131 Karlsruhe, Germany
Tel: +49 721 6084 7734
Email: martin.kagerbauer@kit.edu

Peter Vortisch
Institute for Transport Studies, Karlsruhe Institute of Technology (KIT)
Kaiserstrasse 12, 76131 Karlsruhe, Germany
Tel: +49 721 6084 2255
Email: peter.vortisch@kit.edu

Paper prepared for presentation and publication at the
98th Annual Meeting of the Transportation Research Board
6,762 words
2 tables (2 * 250 words = 500 words)
Total: 7,262 words (including abstract, text, references, cover page, and tables)
Submission date: August 1, 2018

ABSTRACT

With the growing usage of the internet, the possibility for shared mobility has risen just as much. Beside ride-sharing, bike-sharing, and shared parking, this applies, especially to car-sharing.

Past research activities have often been limited to the economic, ecological, and urban benefits of car-sharing, such as the number of privately owned cars that could be replaced by car-sharing vehicles or the potential to save parking space. These analyses disregard the user's behavior and patterns of usage. However, to analyze, e.g., future market shares of car-sharing, we first have to evaluate how car-sharing members use car-sharing and what purposes the trips might serve. One such study has been conducted in Germany, however, using free-floating car-sharing data.

Our focus of research is put on data from a station-based car-sharing provider and what kind of user or usage profiles can be identified. We investigated this by performing a cluster analysis using the k-means algorithm. The results indicate that there are five types of station-based car-sharing users and usage respectively. There are commercial users, users who use car-sharing for regular and users who use it for irregular activities. Furthermore, car-sharing vehicles are used to replace a second car and also for long distance travels.

These findings are in part consistent with the study on free-floating car-sharing but also show some dissimilarities, as to be expected since the two systems generally serve different purposes.

INTRODUCTION

Due to the increased connectivity amongst users through internet access, the possibility to share goods has never been simpler. In transportation, sharing systems like ride-sharing, bike-sharing, and car-sharing emerged. Car-sharing offers the possibility to have access to a variety of vehicle types at a specific time without having to own a car. Insurance and fees are included in a yearly membership fee. The only accruing costs are variable charges for hours rented and distances traveled. Organizational expenses are incurred at the beginning of the membership, making the actual use of car-sharing vehicles much easier than renting cars.

There are two types of car-sharing: free-floating and station-based. In a free-floating system the users pick up a car at its current location and can park it anywhere, usually within a given radius of action. Station-based car-sharing users can pick up cars at a specific location and have to return it to this location as well. The main advantage of station-based car sharing is that a journey can be planned a long time ahead and the user has certainty that the car is available for the reserved period of time. While users of free-floating car sharing cannot plan ahead of time and may have to switch to a different mode of transport when a car is not in reach, they are much more flexible when it comes to booking and parking the car, making this system very attractive for short one-way trips.

Users of both station-based and free-floating car sharing are prone to not own a private car and are more likely to have commuter tickets for public transportation and bicycles. Furthermore, cycling and traveling by public transport are chosen more often in everyday mobility (*3; 4; 5; 6; 7*). This effect is especially noticeable for routine trips like commuting to work or places of education (*1*). Furthermore, both their intermodality, and multimodality are more distinctive than the one of non-car-sharing members (*2*).

Previous research suggests that mainly the financial, ecological, and urban benefits of car-sharing have been investigated. Cities often limit their analyses to the number cars that could be replaced by shared vehicles, the public space gained by this development, and the resulting shifts from car use to use of other modes of transport. Many studies observe the sociodemographic and attitudinal background of car-sharing users, i.e., the sociodemographic groups that are the early adopters, their preferred modes other than car, and their behavior in general. Literature review shows multiple studies addressing the user characteristics of car-sharing. However, there has been little research on car-sharing usage itself (*3*). Investigation of usage patterns of both free-floating and station-based car-sharing is still in its early stages.

This paper aims to delineate usage profiles of station-based car-sharing. To define usage profiles of station-based car-sharing, we have conducted a cluster analysis on bookings made by users, allowing for interpretation of usage attributes, for instance, to differentiate between everyday and holiday trips. The analysis furthermore allows for a more in-depth understanding of the car-sharing market, and to gain insights on what kind of trips are currently conducted using car-sharing vehicles. The results are compared to a similar study in which data from a free-floating car-sharing system were the basis of cluster analysis.

Following the introduction, we present an overview of research activities and the respective literature. The subsequent section of the paper gives insight into the used dataset, followed by a descriptive data analysis. After presenting results from the preliminary data analysis, we go on to elucidate the clustering method we used for our classification. Subsequently, the results are presented, and the clusters are described. Afterward, the presented results are compared with past findings in this research area.

LITERATURE REVIEW

Overview of car-sharing

Multiple studies of data from different cities have already observed the usage of car sharing on a macroscopic scale showing differences in usage depending on system and city. On average, car-sharing members of station-based systems conduct between 12 and 20 trips per year ranging to an average distance of about 42 kilometers whereby distances travelled by business members are about 10% longer. Average booking time amounts to six to seven hours (*4; 5; 1*). The distribution of trips over the week reveals a similar periodicity for both car-sharing trips and private car trips. Even though users of both station-based and free-floating car-sharing show higher frequencies of usage on the weekend, this is more prevalent for station-based car-sharing users. On the other hand, free-floating car-sharing users tend to make trips later in the day (*6; 5*). Distances of car-sharing trips vary significantly from city to city. This might stem from the density of car-sharing stations and the resulting longer time it takes to access a station. In Mannheim, 70 to 80% of car-sharing trips starting in the city are also conducted within the city (*7*) whereas 53% of car-sharing trips in Dresden exceed 50 kilometers leaving the city districts (*8*).

Days, on which car-sharing is used, differ considerably from days where car-sharing is not used. Users are more mobile regarding total distances traveled and number of trips conducted when using car-sharing (*7*). Car-sharing is often used for transporting goods or people and for leisure trips, where a certain degree of flexibility and convenience is needed (*5; 7; 9;*). This applies to station-based car-sharing as well as to free-floating car-sharing. However, these usage patterns are revealed more strongly for station-based car-sharing, whereas free-floating car-sharing is sometimes also used for commuting trips or trips to the airport (*9; 5*). Due to usage patterns, the average occupancy of station-based car-sharing vehicles was significantly higher than of free-floating car-sharing vehicles or privately owned cars (*9*).

Car-sharing vehicles are also used for holiday travel resulting in using car-sharing for long-distance trips (*8*). In this case, the trip is usually dependent on the availability of a (certain) vehicle, leading to the fact that trips may be postponed if the car-sharing vehicle is not available at the preferred time. To avoid having to adapt holiday trips, users often reserve their desired vehicle a long time in advance.

As trips conducted with free-floating car-sharing vehicles are more spontaneous, unavailability of a vehicle results in users having to choose a compensatory mode of transportation immediately, e.g. public transportation (*9*). Furthermore, participants reported taking fewer spontaneous trips in general as a result of a growing awareness of travel costs (*9; 10*).

Most station-based car-sharing users state to have a preferred station located close to their place of residence, mostly within the range of 500 meters. Therefore the access to the station happens mainly on foot or by bike (*7*).

Clustering in transportation research

To group similar typologies and characteristics of individuals or transport users, clustering has shown to be a proven statistical method. Von Behren et al. (*11*) combined German survey data on travel behavior and attitudes towards multiple transport modes, social norms, and preferences to identify different urban mobility types. A similar approach was chosen by Steding (*12*) clustering orientation and attitudes towards private cars, busses, and bicycles.

Regarding car-sharing, mainly characteristics and attitudes of individuals have been the scope of previous studies. Braun et al. (*13*) developed a multiple regression model to investigate the sociodemographic characteristics of car-sharing users, and to analyze the effects of structure (density, access to public transportation, parking, land use). Schreiber et al. (*1*) clustered car-sharing users based on stated attitudes towards different modes of transportation and compared the results with those of a control group. The findings revealed that car-sharing users are mostly represented in bicycle and public transport affine clusters (*1*). A similar cluster was developed by Schreier et al. (*5*). Based on stated attitudes towards car-sharing, public transportation, and private automobile users were differentiated by affinity for car-sharing, affinity for public transportation and grade of rationality/emotionality in mobility. The findings show clusters being addressed by different car-sharing systems. People less affine to public transportation are more attracted to a free-floating system.

To look at usage typologies of car-sharing it is necessary to take behavior on an individual level into consideration. Such typologies of free-floating car-sharing members were investigated by Harz (*3*). Car-sharing members were clustered into five clusters depending on trip frequency, a preferred usage during night time or on weekends, the regularity of trips and the duration of trips. As literature points out, usage patterns of free-floating and station-based car-sharing vary significantly.

DATA

Data preparation

The data source used for the analysis consists of bookings of the most significant car-sharing provider located in Karlsruhe. In 2012, about 5% of households in Karlsruhe had a membership for the aforementioned company (*14*). Regarding the number of stations compared to the number of inhabitants, Karlsruhe is the leading city concerning car-sharing in Germany with a ratio of 2.71 car-sharing vehicles per 1000 inhabitants, followed by Stuttgart (1.47), Freiburg (1.41), and Cologne (1.27) (*15*). Stadtmobil, a company founded in 1999, provides a station-based car-sharing system in both the city and in the regional area of Karlsruhe. The 277 stations are spread over 25 cities and villages in the region of Karlsruhe, as shown in Figure 1. Even though car-sharing is widespread and the quality of supply is relatively high compared to other German cities, the general share of car-sharing trips and therefore its relevance remains low compared to other modes.

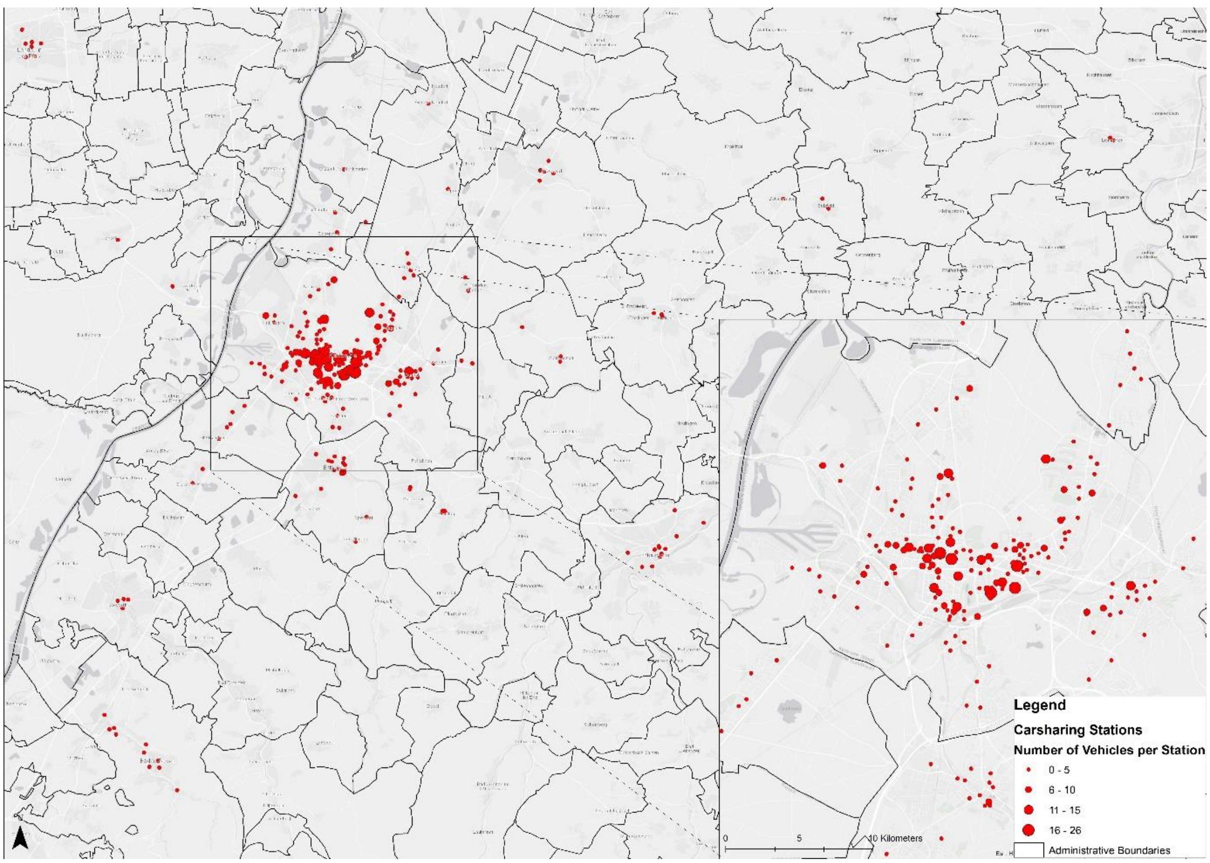

Figure 1: Car-sharing stations in the region of Karlsruhe

The 944 cars are differentiated into various tariff classes: class ”Mini” represents cars with the lowest charge and the smallest size, class “Transport” raises the highest charge and offers the largest vehicles with better equipment. Bookings of the classes “Maxi” and “Transport” are not contained in the used dataset. In each class, a member can choose between various car brands and models. Overall thirteen car brands and 52 models can be booked. Table 1 shows the assignment of car manufacturers and their respective models to tariff classes. It should be mentioned that vehicles are not categorized solely by size but features and prestige of the brand. For example, an Audi A1 is comparable to an Opel Corsa regarding size, however compared to Opel, Audi is considerably more expensive and the A1 is therefore categorized into the “Medium” tariff class over the “Small” tariff class.

Table 1: Assignment of car models to tariff classes

Manufacturer	**Tariff Classes**					
	Mini (A)	**Smart (B)**	**Small (C)**	**Medium (D)**	**Station wagon (E)**	**Large (D)**
Audi				A1	A3	A3 A4
BMW					1 series i3*	2 series 3 series
Fiat		500				
Ford			Fiesta		Focus	
Mazda				MX5		
Mercedes					A-Class	
Mini				Cooper Roadster		
Opel		Karl	Corsa		Ampera Astra	Insignia Zafira
Renault		Twingo	Zoe*	Captur Kangoo Fluence*		
Seat	Mii				Ateca	
Smart		Smart Smart*				
Toyota	Aygo	Yaris		Auris		
VW		UP	Polo	Caddy Golf Golf*	Beetle Jetta	Passat Tiguan

* *electric vehicles*

The dataset contains all bookings of the year 2017, including all trips started between 1st January 2017 and 31st December 2017. This sums up to 232,675 bookings in total. Each booking provides a unique booking id and a unique member id for each member. The bookings are differentiated by type of member into the categories “private” and “business”. There are 201,985 private and 30,690 business bookings. Furthermore, each booking contains the time of reservation, the start and the end of the booking period. The time of reservation denotes the latest time the booking was changed. Therefore, it is possible that this time is within the booking period, defined by the start and end of the booking during which the car was reserved for the user.

Next, the start and the end of a trip is given for each booking. Those are the times when the car is taken until it is brought back by the member. The distance driven by the member is also contained. The station where the car belongs to is logged together with the model and the tariff class.

All bookings were conducted by 13,108 members. Additionally, there are 2,215 registered users who did not book a car in 2017. The dataset contains the gender and age. The gender is available as male, female and other. Male and female members are single person members, while the others are groups of people. The car-sharing company allows partners of members to register as "partner". In this case, the membership costs and registration fees are lower. There is also an affiliate program. In this case, the membership costs depend on the number of registered users per affiliate program. In both cases, the gender of each member of the group is denoted as other, while the age is not available.

Analyzing the distance driven in total reveals 153 million km. The smallest distance is 1km, and the largest is 3.4 million km. This is too high to be driven. Therefore, the bookings are validated and filtered based on the driven distance and the booking speed. The booking speed is calculated by the distance driven relative to the duration of the trip. Comparable studies use the differentiation between driving mode and parking mode to calculate the driving time (*6*). The existing data does not allow for a distinction between a parked and a driving car. Therefore, the speed used to make bookings plausible is not called travel speed. The driven distance is limited to 5,000 km and the booking speed to 50 km/h.

After filtering the bookings, 227,624 bookings remain in the dataset. 197,885 done by private and 29,739 by business members. The total distance of all bookings is 20 million km, which are traveled by 13,027 members. After validation of the dataset, the smallest distance remains at 1km and the largest is lowered to 4,886km.

Of the remaining shares of bookings, 49.7% are done by male and 30.3% by female members. For the remaining 20.0% the gender is other. 78.4% of the members contain information about their age. Most of the bookings with missing age or gender are made by business members. Unfortunately, there are entries available, where only the gender or the age is available. Overall, only 77.7% have both the correct age and gender. Due to the inconsistent information about the gender and age, those two attributes are ignored for further analyses.

The time of the reservation is also not considered, because it is updated on each change. Due to this, it is not possible to analyze the buffer time between the reservation and the real start of the trip. One of the changes, which could be done, is changing the start or end time of the booking. Due to this, it is also not possible to analyze the difference between the initially reserved booking period and the needed period by the trip. Therefore, the booking period is not used in the analyses.

Based on this, the following attributes remain and are used in the analyses: start time, end time, duration and travel distance of each trip, tariff class, and member id.

Descriptive Analysis

Members rent on average 17.5 car-sharing cars per year, ranging from only one booking to 648 bookings. The cars are rented on average about 10 hours. The shortest booking was 2 minutes, while the longest was 53 days.

The bookings are widespread over all ages. The youngest member is eighteen years old, which corresponds to the minimum driving age in Germany. The oldest one is 89 years old. The number of bookings per age corresponds more or less with the number of registered users per age, as shown in Figure 2. It can be seen, that most of the users are between 27 and 55 years old. The number of bookings per customer per age is more or less the same over all ages.

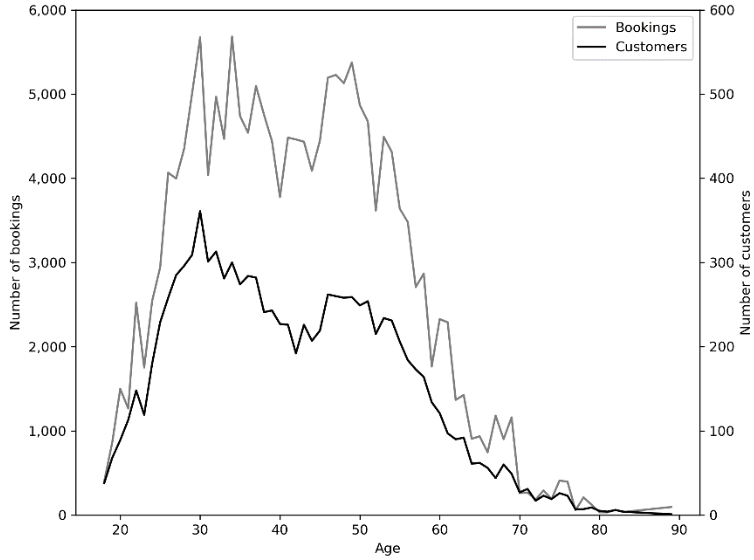

Figure 2: Number of Bookings and members grouped by age

Taking a look at the different tariff classes and cars per tariff class results in a plot as presented in Figure 3. This plot shows that there are more bookings per car for tariff classes “Mini”, “Smart”, and “Small” compared to tariff classes “Medium”, “Station wagon”, and “Large”. This comparison corresponds to the average and median duration and distance of trips done with the different tariff classes. Furthermore, tariff classes “Mini”, “Smart”, and “Small” are used for shorter trips, while tariff classes “Medium”, “Station wagon”, and “Large” are used for longer trips. Mapping tariff classes to vehicle size shows that “Mini”, “Smart”, and “Small” contain mainly small to medium cars, while tariff classes “Medium”, “Station wagon”, and “Large” contain medium to large cars. The most often booked tariff class is “Small”, where most of the vehicles are available. As class “Small” also contains medium-size cars, those cars are used for both, short distance and long distance trips.

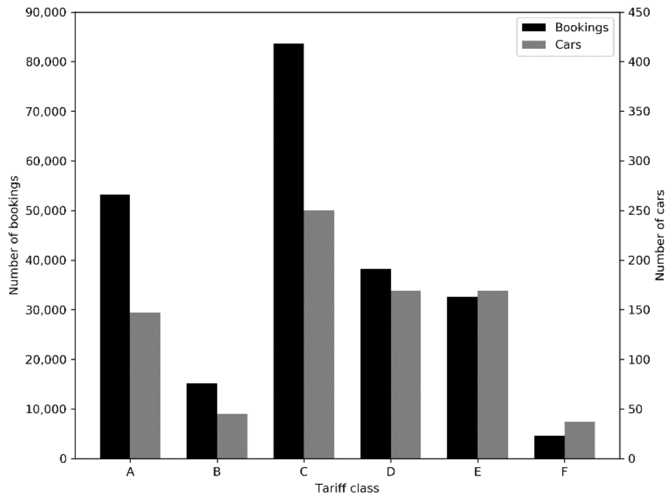

Figure 3: Number of Bookings and cars grouped by tariff class

Given the described dataset, it is not possible to distinguish between individual cars. It is only possible to analyze the trips per tariff class and station and normalize this by the number of available cars at the corresponding station. This results in an average of 246 trips per car during the year or 0.68 trips per day. Figure 4 shows the distribution of average trips per day and car. It can be seen, that most of the cars are used on average between 0.25 and 1.6 times a day. There are only a little number of cars used more than 1.6 times a day.

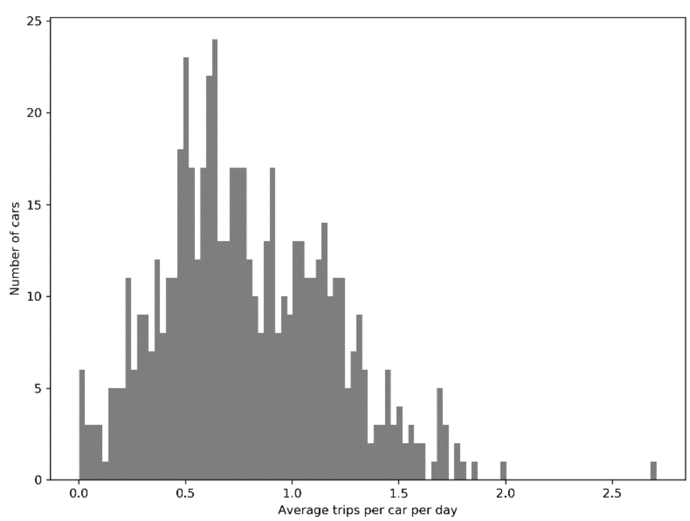

Figure 4: Average trips per class and station normalized by the number of cars

Based on the overall duration of all trips and the number of cars, the average utilization per car is 31.4%. This corresponds with the distribution shown in Figure 5. It shows the number of cars for each average utilization per car. The utilization per car is based on the utilization per tariff class and station normalized by the number of cars at the corresponding station. As can be seen in Figure 5 most of the cars are used between 10% and 50% of the day.

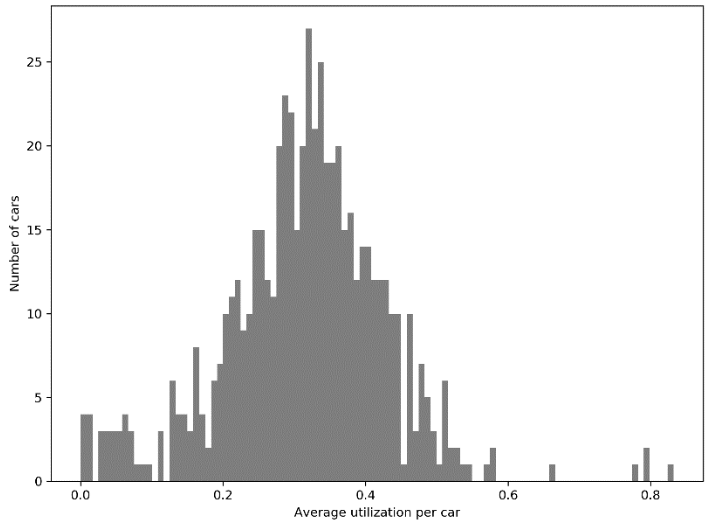

Figure 5: Utilization per class and station normalized by the number of cars

CLUSTER ANALYSIS

A common practice in explorative data analyses are clustering methods. The primary goal of cluster analysis is to group a large set of objects to find an empirical classification. Probably the most common used algorithm for cluster analysis is the k-means algorithm. It is a deterministic clustering method that groups n objects into k clusters by minimizing the variance. This algorithm should not be applied when clusters overlap and the sample size is very small. In case of nearly no overlap, medium sample sizes are acceptable. If only a small sample size is available, other clustering algorithms should be applied, such as the ward method, as k-means would lead to unstable results. As our sample size is sufficiently large especially regarding the small number of variables, we chose to apply the k-means algorithm to our data. (*16*)

The cluster analysis was implemented using the programming language R. R provides several packages for cluster analysis. For our analysis, we used the package *factoextra* to determine the number of possible clusters and the package *stat* for the actual cluster analysis.

Choosing variables for cluster analysis

As aforementioned, the data is available as booking data. As the intended goal was to cluster members and not individual bookings, we first grouped the data by member-id. For each characteristic, either the median or variance was calculated and used as input for the clustering. For preliminary data analysis, we calculated the correlation matrix for the variables of interest. The variables we identified to influence user behavior and usage profiles, and therefore the clusters are:

1. Frequency of usage
2. The distance of single trips
3. The variance of trip distances
4. Size of the vehicle/tariff classes (mini, smart, small, medium, station wagon , large)

5. Day of the week the trip started

The dataset only contains basic membership data and does not allow for in depth description of the users and their specific trip purpose. Therefore, the information on a membership being classified as private or business is only suitable for differentiation between pricing schemes. We chose to exclude the variable business/private from the cluster analyses due to missing information on whether the members used the cars solely for business or private purposes respectively.

The analysis of the correlation matrix showed that there is a moderate relationship between the frequency of car-sharing use and some of the tariff classes and the days Monday through Friday. Furthermore, a relationship between the tariff classes “Smart”, “Small”, and “Medium” and weekdays, and between the weekdays among each other can be identified. Consequently, the next step was to decide whether or not to remove correlated variables. The correlation mainly affects tariff classes and weekdays. However, information on the course of bookings throughout the week and the chosen tariff class is quite important when wanting to analyze user behavior. In their work, Mingoti and Lima furthermore present findings showing that the correlation between variables does not significantly affect the result of all clustering algorithms, including k-means (*17*). Therefore, all variables were kept regardless of the correlation relationship between the variables.

Determining the number of clusters

In the next step, the clustering algorithm is applied to the data while simultaneously the number of clusters is determined. This is achieved by applying a statistical method to determine the number of clusters, which is then verified by a heuristic method. The statistical method for determining the number of clusters is called the “gap statistic”, where the scattering of a reference null distribution is compared to the change in within-cluster scattering (*18*). The result of the gap statistic calculations is shown in Figure 6. The plot not only allows for the selection of a suitable number of clusters but also for an evaluation on whether or not cluster analysis should be applied in the first place. As indicated in the graph, the one-cluster solution is not feasible. Therefore the application of cluster analysis is sensible. The plot shows three possible number of clusters. The first break is at two clusters, however, the gap function increases after five and again after 12 clusters, which should also be taken into consideration for the possible number of clusters.

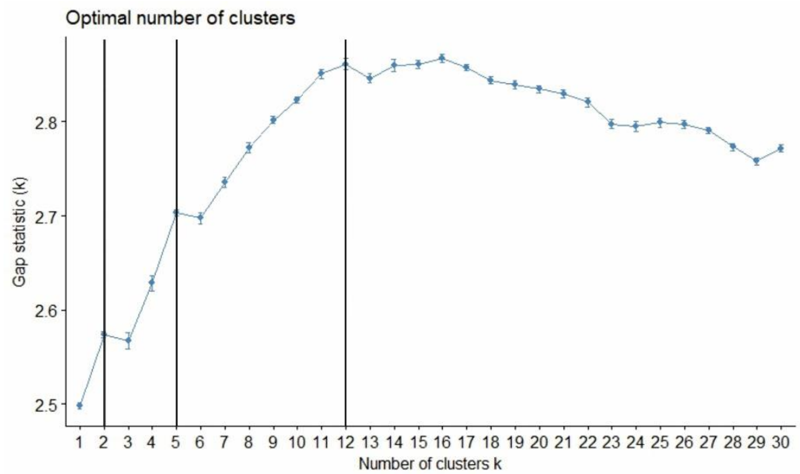

Figure 6: Result of gap statistic analysis

A heuristic approach to determining the right number of clusters is the elbow method, where the total within sum of squares is plotted against the number of clusters. The optimal number of clusters is subsequently identified by an "elbow" in the curve. The elbow-plot for the car-sharing data is shown in Figure 7. The plot shows that the optimal number of clusters is either two or five.

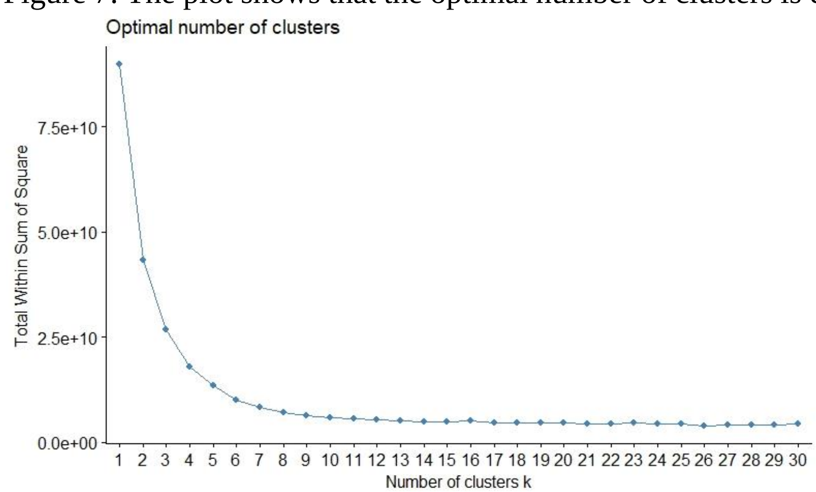

Figure 7: Elbow-plot for identifying the optimal number of clusters

Results

Following the step of determining the number of clusters, the k-means algorithm was applied to the data for both two and five clusters. The results of the two clustering analyses are presented in Table 2.

Table 2: Results of cluster analyses

Variable	5-Cluster solution					2-Cluster solution	
	A	B	C	D	E	A	B
n	9	1928	6783	329	1022	9052	1019
Median Frequency	408	45	9	114	7	11	74
Median Trip Distance [km]	17	26	38	23	279	42	23
Median Trip Distance Variance [km]	110	82	66	62	443	82	71
Median Trips in Tariff Class Mini	48	6	1	27	0	1	14
Median Trips in Tariff Class Smart	22	0	0	1	0	0	1
Median Trips in Tariff Class Small	111	15	3	38	2	3	25
Median Trips in Tariff Class Medium	69	6	1	10	1	1	9
Median Trips in Tariff Class Station Wagon	140	4	1	6	2	1	5
Median Trips in Tariff Class Large	11	0	0	0	0	0	0
Median Trips on Mondays	92	5	1	16	1	1	10
Median Trips on Tuesdays	83	5	1	19	1	1	10
Median Trips on Wednesdays	76	5	1	18	1	1	10
Median Trips on Thursdays	75	6	1	20	1	1	11
Median Trips on Fridays	47	7	1	19	1	2	12
Median Trips on Saturdays	0	8	1	11	1	2	10
Median Trips on Sundays	4	5	1	6	0	1	6

The presented 2-cluster solution does not reveal any obvious user or usage profiles. Cluster 1 (A) contains 9,052 members, while the remaining 1,019 members are included in cluster 2 (B). The members are divided between more and less frequent users. Members conducting fewer trips tend to take longer trips and vice versa. The results of the 2-cluster method show no indication of specific usage profiles. Due to this observation, the 5-cluster solution was chosen to be analyzed in detail.

The 5-cluster solution allows for detailed analysis and identification of the user, respectively usage profiles. The findings regarding median trip distance and median trip frequency correspond to findings of the 2-cluster solution as presented in Figure 8, in which data points for both trip distances and trip frequencies are plotted for each cluster. Trip distance decreases with increasing trip frequency and vice versa.

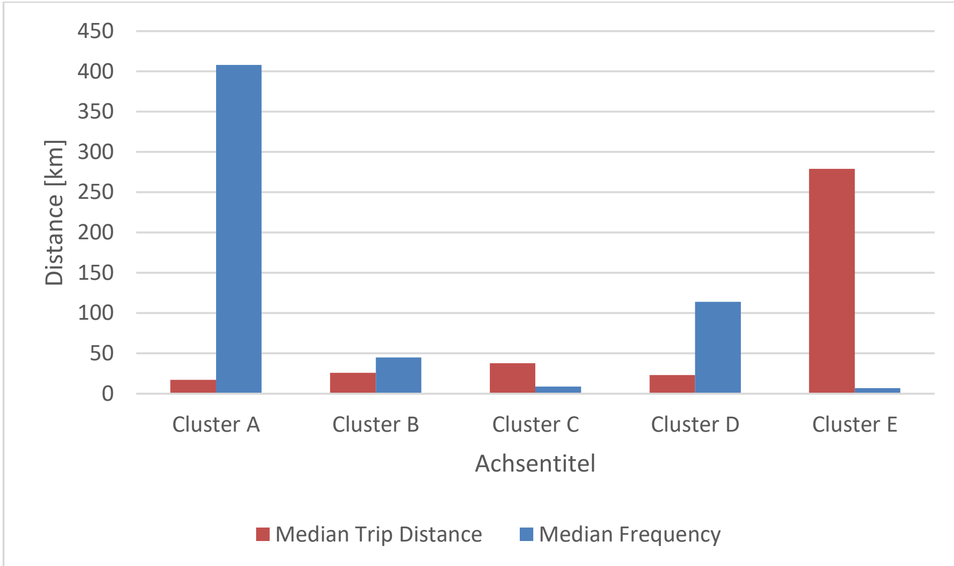

Figure 8: Median trip distance and frequency of the different clusters

Figure 9 presents an overview of the used tariff classes and consequently the vehicle size. The radar plot shows the percentage share of the vehicle sizes for each cluster. The plot illustrates that there are three clusters that tend towards smaller vehicles, while the other two are more drawn to larger vehicles. The members of clusters B, C, and D all tend toward mainly using the tariff classes mini and small. Members of cluster E only use vehicles of tariff class mini occasionally. They prefer vehicles of the tariff classes small, medium, and especially station wagon. This is also the only cluster where members use vehicles from tariff class large, however, even in this cluster, it only applies to a small percentage. The plot for cluster A shows a similar shape. However, the members do not use vehicles from tariff class mini at all, as their trips are solely conducted with vehicles from tariff classes smart, medium, and station wagon.

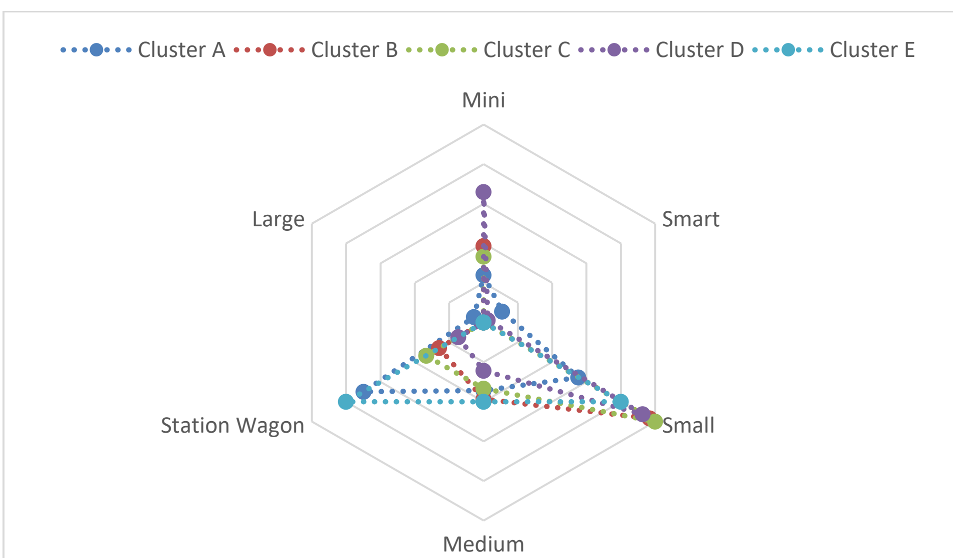

Figure 9: Radar plot of the share of used tariff classes in the five clusters

The distribution of trips by weekdays is presented in Figure 10 in which the values represent the shares of the weekdays for each cluster. This plot shows how stable the trips are throughout the week. Members of cluster C achieve the evenest distribution of trips. Members of cluster B tend to be more active towards the weekend, while members of clusters A, D, and E reduce their trips on the weekend, especially on Sundays.

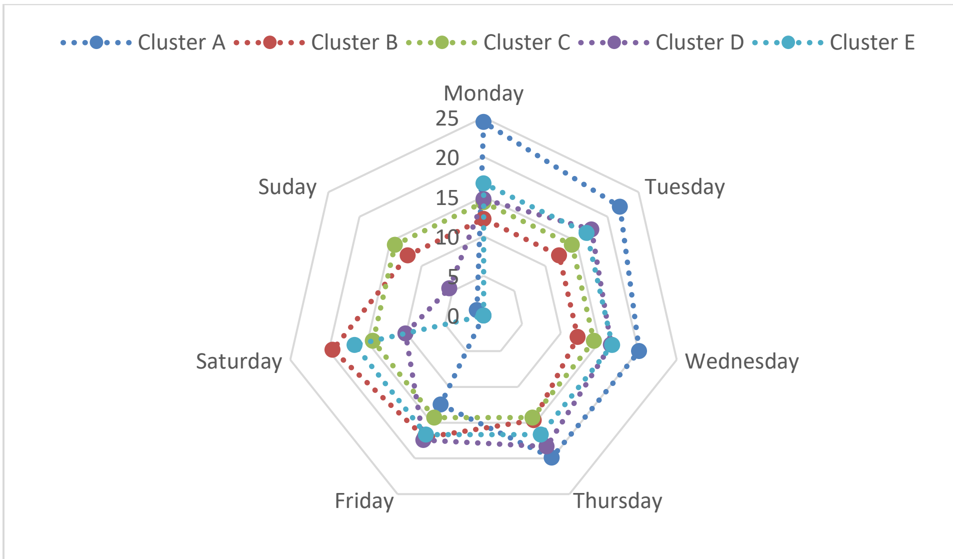

Figure 10: Radar plot of the number of trips during different weekdays for the five clusters

Each cluster presents key characteristics, which are further discussed in the following section. The traits of the clusters allow for classification and identification of user and usage profiles respectively. However, as the data does contain trip purposes, these are only elementary descriptions and the names and specifications of the clusters have to be verified in future studies.

Commercial Users (A)

Cluster A is composed of nine members and therefore the smallest cluster. It contains members who use car-sharing frequently, on average more than once a day. They are defined by very short trip distances. However, the variance of trip distance is very high. The users tend to use small vehicles or station wagons. The number of trips during weekdays is stable. Trips on the weekend are rather infrequent. Due to this behavior, we identified these members as commercial users. On average they use car-sharing vehicles just over once a day, indicating that those are members that have access to more than one vehicle at a time. This applies to businesses, which tend to have more cards to access car-sharing vehicles simultaneously.

The findings suggest that these users would benefit more from a free-floating car-sharing system, as the trip distances are small. However, using cars for business requires a certain degree of certainty that there will be a car for the return trip, which is given when using station-based car-sharing.

Regular Weekly Activities (B)

Cluster B contains 1,928 members and is defined by a medium trip frequency and rather short distances, while the trip distances vary only a little. The tariff class small is the one most frequently used, while the classes mini, medium, and station wagon are occasionally used. Usage during the week is somewhat stable, while the bookings increase towards the end of the week. The trip frequency indicates that the car-sharing vehicles are used once a week to do errands or go grocery shopping. A survey conducted in Germany on the preferred days of grocery shopping clearly indicates that the primary weekdays for grocery shopping are Friday and Saturday, complying with the trips of the members in this cluster (*19*). The regular trips on Sundays can be explained by regular weekly leisure activities such as sports events.

Irregular Activities (C)

Cluster C is the largest cluster and contains 6,783 members who only use car-sharing occasionally as indicated by the small trip frequency. Concerning used tariff classes, cluster C is similar to cluster B. Half of the bookings are in tariff class small, while the rest is equally distributed among tariff classes mini, medium, and station wagon. Throughout the week, the bookings remain entirely stable. We have identified this group as one applying to members using car-sharing vehicles only on occasions where using a car is inevitable. This applies to trips where, e.g., equipment needs to be transported, or there is no public transport access.

Substitution of public transport trips can primarily be observed in free-floating car-sharing trips, however, during the time of the data collection, there was no free-floating system implemented in the spatial area of study. A lot of the members of this cluster will probably benefit from the recently introduced free-floating car-sharing system, however, further research is needed to verify these observations.

Second Car Replacement (D)

The second smallest cluster is the fourth cluster, containing 329 members. Those members use car-sharing regularly, on average about three times a week. The trips themselves are rather short, and the distances vary only a little. The majority of trips are conducted using small vehicles. About 80% of all bookings by members in cluster D are executed using vehicles of tariff class mini, smart or small. The number of bookings during weekdays is stable and decreases towards the weekend. We identified these members as users that replace a second car in the household with a car-sharing vehicle. They regularly conduct trips during the week, such as doing errands or picking up children. The primary car is likely to be used for trips to and from work by the main earner during the week. On the weekend the trips conducted by the car-sharing vehicles decrease as the primary car is available for errands.

The identified trips are primarily round-trip and the time that the car is not in use, i.e. driving, is limited. Therefore, the station-based car-sharing system is suitable for the users in this cluster and the members of cluster D benefit from the pricing scheme.

Travelers (E)

The final cluster of the 5-Cluster solution contains 1,022 members and is defined by very long but rare trips. The first two tariff classes are not used indicating that larger cars are preferred for the trips. The number of trips is rather stable during the week but tend to increase towards the weekend. These results indicate that the cluster includes members using car-sharing mainly for traveling purposes. The median trip distance is much longer than in the other clusters. However, also the variance is much higher than in the remaining clusters. This can be explained by the fact that holiday trips differ a lot in distance. Furthermore, members using car-sharing for primarily for travels also take advantage of being a car-sharing member, using the vehicles for different purposes on occasion. This is supported by the trip frequency: holiday trips are usually conducted two to three times a year. The rest of the trips taken with a car-sharing vehicle serve a different purpose, which tend to be much shorter.

Comparison to similar study

As mentioned in the literature review, there has been a similar study, however, focusing on free-floating car-sharing users. Due to the attributes of free-floating car-sharing systems, users tend to behave differently than users of station-based car-sharing, however, the work presented by Harz (*3*) still reveals some overlap of the clusters. He differentiates between five clusters:

1. Seldom Users
2. Users with long trip times
3. Standard Users
4. Frequent Users – Commuters
5. Frequent Users – Night Trips

The characteristics of cluster 1 correspond with those of cluster C from this study. Both analyses find that these clusters are defined by members who use car-sharing only sporadically. The clusters contain many users who account for a little share of the trips.

The study of Harz suggests that members in cluster 2 are defined users with long trip times. This cluster corresponds to cluster D of our study, which contains members who use car-sharing vehicles as a second car replacement. Harz suggests that members of this cluster conduct more trips during one booking, indicating several stops and therefore multiple activity purposes. This corresponds to the usage of a station-based car-sharing vehicle for the purpose of doing errands, going grocery shopping or picking up children.

Cluster 4 of Harz's clustering analysis includes members that use free-floating car-sharing for regular commuting trips. Even though some characteristics are comparable to those of cluster A, such as above average trips on weekdays and a small user group with a large share of trips, the trips conducted with the car-sharing vehicle probably do not have the same purpose. As free-floating car-sharing allows for the usage of cars on a one-way trip, it can entirely substitute commuting trips using public transit. Due to the time-dependent costs and the fact that cars need to be returned to the station, these trips are usually not replaced by station-based car-sharing vehicles.

The two other clusters show no indication of overlap. However, it was to be expected that there would not be a complete overlap due to the fact that free-floating and station-based car-sharing each show different advantages and disadvantages, which are reflected in the respective datasets.

CONCLUSIONS

People owning a car are prone to use it more often and are not as likely to switch to another transport mode than people not owning a private car. Car-sharing can be used to substitute private cars and make cities more liveable. However, for cities to actually estimate the impact of introducing a car-sharing system, more insight into the usage profiles of car sharing members is needed.

In order to make car sharing usage more accessible, we conducted cluster analyses. We have chosen the described approach to classify our data based on interpretability of results and the comprehensible algorithms used. This allows for a transparent classification process, however, due to the well-known limitations of cluster analysis, we suggest that more complex classification algorithms should be applied to gain more insight into the data. The analyses are based on a dataset containing roughly 220,000 individual bookings of about 13,000 members. This is sufficient to extract usage profiles for different kind of car-sharing users. Using k-means clustering, 5 clusters were identified – Commercial Users, Regular Weekly Activities, Irregular Activities, Second Car Replacement and Travellers. Those clusters are comparable to the ones found by Harz (*3*), who used data from a free-floating car-sharing system to analyze usage behavior.

Our analyses allow for identification of typical usage scenarios helping cities and car-sharing providers to match their supply to the actual demand. Furthermore, our results help to determine in which scenarios use of a station-based car-sharing vehicle is not attractive, and therefore show how e.g. the pricing system can be adapted to be appealing to more users.

As the results can be compared to other studies, the methodology should be transferable to other data from other cities as well. When it is applied to data from another city, we strongly recommend analyzing the available attributes for each booking. If available, the analysis can benefit from information about individual cars. Based on this, one can analyze preferences to cars and also analyze the utilization of cars better. To gain more insight into the behavior of car-sharing users, we recommend querying the purpose of a trip and detailed personal characteristics. This would allow for verification of the identified clusters.

ACKNOWLEDGMENTS

The data was provided by Stadtmobil Car-sharing GmbH & Co. KG as part of the regiomove project, funded by the State of Baden-Württemberg and the European Regional Development Fund.

AUTHORS CONFIRMATION OF CONTRIBUTION

The authors confirm the contribution to the paper as follows:

- Literature review: Tim Wörle, Anna Reiffer
- Data preparation: Anna Reiffer, Lars Briem, Tamer Soylu
- Data analysis: Lars Briem, Anna Reiffer, Tim Wörle
- Cluster analysis: Anna Reiffer
- Interpretation of results: Anna Reiffer, Lars Briem, Tim Wörle, Martin Kagerbauer, Peter Vortisch
- Draft manuscript preparation: Tim Wörle

All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. Schreier, H., C. Grimm, U. Kurz, B. Schwieger, S. Keßler, and G. Möser. Analyse der Auswirkungen des Car-Sharing in Bremen, 2018.
2. Kopp, J., R. Gerike, and K. W. Axhausen. Do sharing people behave differently? An empirical evaluation of the distinctive mobility patterns of free-floating car-sharing members. *Transportation*, Vol. 42, No. 3, 2015, pp. 449–469.
3. Harz, J. Variablen-Verdichtung und Clustern von Big Data - Wie lassen sich die Free-Floating-Carsharing-Nutzer typisieren? *Verkehrsökologische Schriftenreihe, No.* 5, 2016.
4. Haefeli, U., D. Matti, C. Schreyer, and M. Maibach. *Evaluation Car-Sharing. Schlussbereicht*, 2006. Accessed July 18, 2018.
5. Schreier, H., U. Becker, and J. Heller. *Evaluation CarSharing (EVA-CS) Landeshauptstadt München*, team red; TU Dresden; Omnitrend, 2015. https://www.muenchen-transparent.de/dokumente/3885730.pdf. Accessed July 24, 2018.
6. Schmöller, S., S. Weikl, J. Müller, and K. Bogenberger. Empirical analysis of free-floating carsharing usage: The Munich and Berlin case. *Transportation Research Part C: Emerging Technologies*, Vol. 56, 2015, pp. 34–51.
7. Huwer, U. *Kombinierte Mobilität gestalten: Die Schnittstelle ÖPNV - CarSharing*, 2002.
8. Perner, T., P. Schöne, and H. Brosig. *Das Dresdner Modell. Car-Sharing und ÖPNV*, 2000. Accessed July 19, 2018.
9. Becker, H., F. Ciari, and K. W. Axhausen. Comparing Car-Sharing Schemes in Switzerland: User Groups and Usage Patterns. In *TRB 95th Annual Meeting Compendium of Papers,* Washington, D.C., 2016.
10. Shaheen, S. A., A. P. Cohen, and M. S. Chung. North American Carsharing. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2110, No. 1, 2009, pp. 35–44.
11. von Behren, S., C. Minster, M. Magdolen, B. Chlond, M. Hunecke, and P. Vortisch. Bringing travel behavior and attitudes together: An integrated survey approach for clustering urban mobility types. In *TRB 97th Annual Meeting Compendium of Papers*, Washington, D.C., 2018.
12. Steding, D., and Annett, eds. *Carsharing - sozialinnovativ und kulturell selektiv? Möglichkeiten und Grenzen einer nachhaltigen Mobilität.* Zentrum für Umweltforschung, Münster, 2004.
13. Braun, A., V. Hochschild, and A. Koch. *Intraregionale Unterschiede in der Carsharing-Nachfrage*, Tübingen, 2013.
14. omniphon (Hrsg.). *Mobilitätsverhalten 2012 - Stadt Karlsruhe: Im Auftrag der Stadt Karlsruhe und des Nachbarschaftsverbandes Karlsruhe*.

15. Bundesverband CarSharing e.V. *CarSharing-Städteranking 2017*. https://carsharing.de/sites/default/files/uploads/rangliste_carsharing-staedteranking_2017.pdf.

16. Bacher, J., A. Pöge, and K. Wenzig. *Clusteranalyse. Anwendungsorientierte Einführung in Klassifikationsverfahren.* Oldenbourg, München, 2010.

17. Mingoti, S. A., and J. O. Lima. Comparing SOM neural network with Fuzzy c-means, K-means and traditional hierarchical clustering algorithms. *European Journal of Operational Research*, Vol. 174, No. 3, 2006, pp. 1742–1759.

18. Tibshirani, R., G. Walther, and T. Hastie. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 63, No. 2, 2001, pp. 411–423.

19. Arbeitsgemeinschaft Verbrauchs- und Medienanalyse (VuMA). *Bevölkerung in Deutschland nach bevorzugten Einkaufstagen für Güter des täglichen Bedarfs (Lebensmittel, Getränke usw.) in den Jahren 2013 bis 2017.* https://de.statista.com/statistik/daten/studie/183281/umfrage/bevorzugte-einkaufstage-von-lebensmitteln-und-getraenken/. Accessed August 20, 2018.