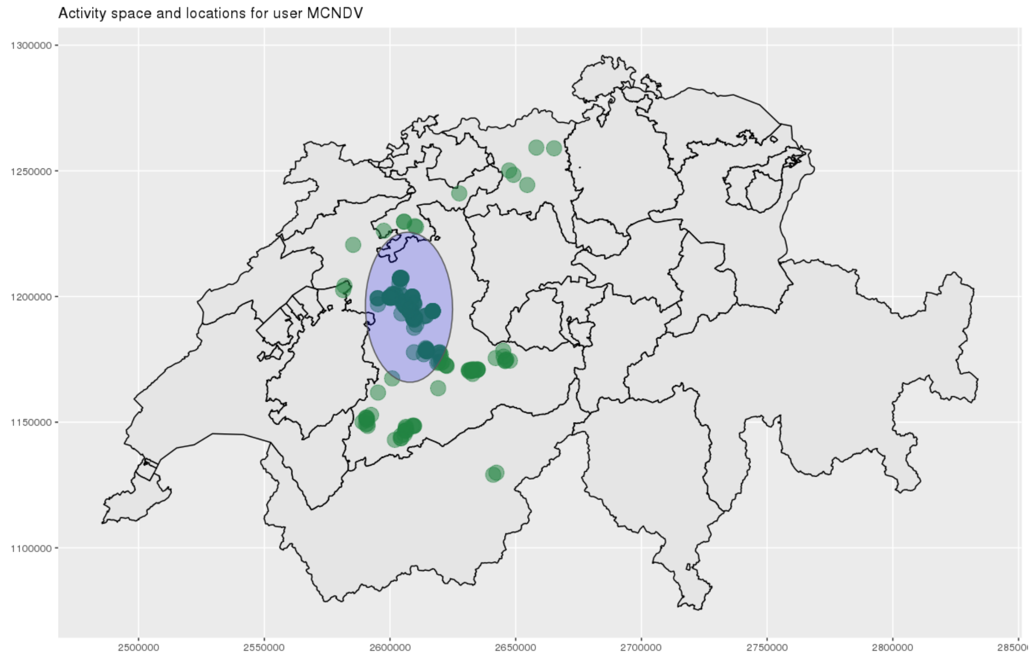


Activity Spaces and Behavioral Innovation

Sara Rytz

Supervisors:

Joseph Molloy and

Prof. Dr. Kay W. Axhausen

MSc Semester project **May 2020**

Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Activity Spaces and Behavioral Innovation

Sara Rytz

IVT
ETH Zürich
CH-8093 Zurich
rytzs@student.ethz.ch

Supervisors:
Joseph Molloy and
Prof. Dr. Kay W. Axhausen
IVT
ETH Zürich
CH-8093 Zurich
joseph.molloy@ivt.baug.ethz.ch
axhausen@ethz.ch

May 2020

Abstract

The goal of this report is to shed light on the socio-demographic aspects of the distribution of activity spaces and innovation rates. Several user attributes such as age, household size, income and their main language are investigated and compared to previous conducted research. Based on the data from MOBIS, individual activity spaces and innovation rates are computed using confidence ellipses and clustering methods. There is no discrepancy between the investigated subsets. Further research is warranted to further evaluate results generated by this paper.

Keywords

Activity spaces, Innovation rate, 95%-confidence ellipses, dbscan clustering, socio-demographic attributes, MOBIS

Suggested Citation

Rytz, S. (2020) Activity spaces and innovational behaviour, Semester project, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zürich, Zürich.

Activity Spaces and Behavioral Innovation

Sara Rytz

IVT
ETH Zürich
CH-8093 Zurich
rytzs@student.ethz.ch

Supervisors:
Joseph Molloy and
Prof. Dr. Kay W. Axhausen
IVT
ETH Zürich
CH-8093 Zurich
joseph.molloy@ivt.baug.ethz.ch
axhausen@ethz.ch

May 2020

Zusammenfassung

Dieser Bericht soll Einblicke in die sozial demografischen Aspekte der Gesellschaft aufzeigen und deren Aktivitätenräume und deren Innovationsraten berechnen. Verschiedene Teilnehmereigenschaften wurden analysiert, darunter das Alter, die Haushaltsgrösse, das Einkommen oder auch welche Sprache die Teilnehmer sprechen. Als Grundlage für neuen Untersuchungen dienen die Daten des MOBIS-Projekts, womit die Aktivitätenräume und die Innovationsraten berechnet werden. Dabei werden die Methoden von Vertrauensellipsen und Cluster-Methoden verwendet. Die Resultate werden mit bereits durchgeführten Studien verglichen und zeigen keine Unterschiede zwischen bestimmten gesellschaftlichen Eigenschaften aus der Teilnehmergruppe. Für eine genauere Beurteilung sind jedoch weitere Untersuchungen nötig.

Schlagworte

Aktivitätenräume, Innovationenrate, 95%-Vertrauensellipse, dbscan Gruppierung, sozialdemografische Eigenschaften, MOBIS

Zitierungsvorschlag

Rytz, S. (2020) Activity spaces and innovational behaviour, Semesterarbeit, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zürich, Zürich.

Contents

List of Tables

List of Figures

1 Introduction

Mobility is crucial to perform daily activities. Throughout evolution mankind traveled and moved in space, whether it was to hunt or for various other reasons. The emergence of new transport systems enabled humans to travel longer distances and to do so faster. Furthermore, the question arises if there are certain groups in society, which perform a different travel pattern than others? In the last few decades various approaches were used to determine the spatial dimensions of individual travelers and research has started investigating those aspects in recent works such as Axhausen *et al.* (2002). However, the lack of sufficient and extensive tracking data has hampered advancements in this area of research.

The data acquisition of the project "MOBIS" has opened a huge variety of new approaches in analysing personal travel behaviour on a large scale data set. Socio-demographic effects on activity spaces and innovation can be considered as confounders of differences in various groups and should therefore help to improve and plan the future of the transportation system in Switzerland. Therefore, this new dataset enables tremendous opportunities for further investigation, as in this paper. In this work, the focus is on insights for different ages, household sizes, incomes and languages. Taking those attributes into account, activity spaces and innovation rates are determined, compared and discussed. This allows an insight into behavioral aspects and gives the opportunity to evaluate the demographic usage of the transport systems in Switzerland.

Previous research and tracking data from other studies produce similar results and that certain trends can be detected. However, it does not indicate for any extraordinary differences between the subsets. Notwithstanding the extensive data acquisition of MOBIS, for a deeper interpretation of the results there are still some crucial information such as the well-being and the family background missing.In first section pf this report an overview on previous research is given and their different approaches are depicted. In the Methods section the approaches are described as well as what the MOBIS data stands for. Further, in section 3 the results are listed and last but not least interpreted in the Discussion and conclusion part.

2 Background

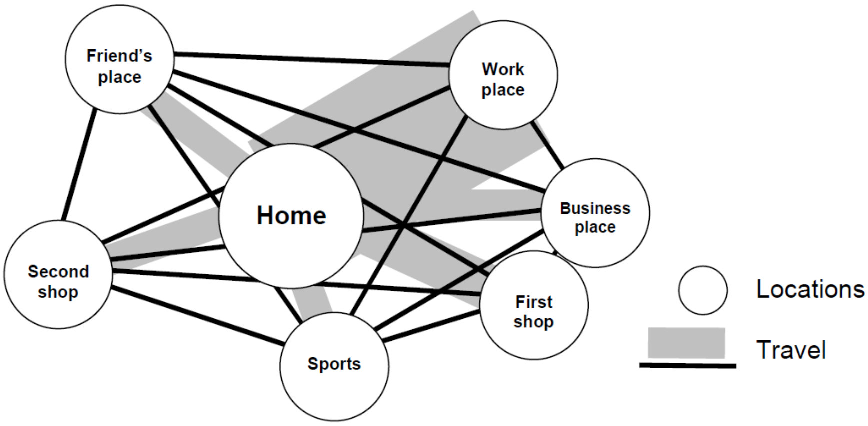
2.1 Activity Spaces

The graphical representation of the spatial environment in which mankind moves has been investigated over the last decades. The first appearance of the space dimensions on human travel behaviour was made in the 1960s and 1970s and was referred to as a traveler's personal world (Lynch, 1960). This so called world was sketched based on a memory protocol of the participants. According to Lynch's work, those personal worlds are biased by the knowledge of the places and the locations the user already knows. Over the years several other concepts have evolved based on the same background and on the same idea of graphically representing a traveler's space. There was no existing term so several wordings were used. These included: awareness space (Brown and Moore, 1970), action space (Horton and Reynolds, 1971), space-time prisms (Lenntorp, 1976), mental maps (Downs and Stea, 1977), perceptual space (Dürr, 1979), activity repertoire or expectation space (Schönfelder and Axhausen, 2003). Whereas the first five spaces stress the potential of travel and their relative positions and connections, the activity repertoire focuses on the type, quality and costs of activities at different locations (Schönfelder and Axhausen, 2003).
Activity spaces in their inherent structure, size and geometry were introduced by Golledge and Stimson (1997). There are three main determinants for an activity space, as they are

- Home
- Regular activities
- Travel between and around these pegs.

Figure 1 shows the idea behind the concept of activity spaces. The importance of a location as well as the connection can be depicted by its size. However, this approach of estimating the size and geometry of the activity space has its limitations due to the available geometric data. In order to predict and estimate a person's environment, precise long-term observations were necessary. So far, the analysed data consists of hand written dairies as in the memory protocols mentioned above (Lynch, 1960). The same methods were applied to the study of Mobidrive (Axhausen *et al.*, 2002). However, in post-processing the activity locations have been given a geocode, so that a graphical representation could be mapped. The emergence of new technologies such as GPS or bluetooth tracking, made

Figure 1: Simplified activity space



Source: Maier *et al.* (1977)

it possible to acquire people's movements digitally and automated. The study of Borlänge was an early example of GPS-tracking as a recording method (Biding and Lind, 2002) (Schönfelder *et al.*, 2002). In these studies geocodes, whether autonomous or manually added, were collected and therefore made it possible to precisely estimate personal activity spaces.

Nowadays, activity spaces aim to represent the distribution and allocation of places. It should depict the visited places and highlight the space in which the persons activities take place. Tracking a person's movements allows to determine the actual travel pattern over a specific time interval (Axhausen, 2002). Additionally, having a longer time frame of observation, helps to improve the stability of the activity space.

Given the new representation of acquired data consisting of geocodes, research has the opportunity to calculate activity spaces for single individuals, whereas earlier only cross-sectional data for groups of respondents were depicted (Schönfelder and Axhausen, 2003). In contrary, the study of Dijst (1999) focused on individual action spaces.

Up to the late 90s, almost no studies regarding calculation and geographical representation of individual activity spaces had been performed. Hence, the limited (only car travel) amount of data in the Borlänge study, the findings show a variability on human activity spaces (Schönfelder and Samaga, 2003). Those findings were further investigated in the Mobidrive study (Axhausen *et al.*, 2002) and also in the work of Axhausen and Schönfelder

(2010). Additional research branches are added such as clustering of activities as well as the interaction of activity density. Computing the actual size of activity spaces, however, was already known. According to Schönfelder and Axhausen (2004) or Axhausen and Schönfelder (2010) three approaches are considered. These are:

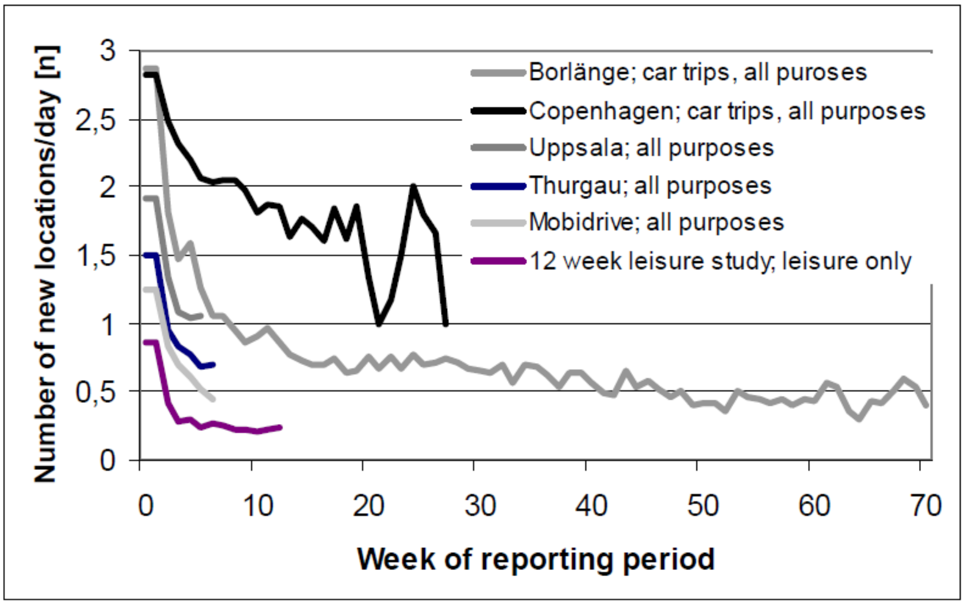
- a two-dimensional confidence ellipse
- kernel densities
- minimum spanning tree (network) or shortest-path network

The kernel densities, as well as the minimum spanning tree are not further elaborated at this point, refer to Axhausen and Schönfelder (2010), Silverman (1986) or Golledge (1999). Probably the most straight forward appraoch are the confidence ellipses, which were introduced in the 1970s UMOT (Unified Mechanism of Travel) (Zahavi, 1979) and follow up in studies such as Beckmann *et al.* (1983a) and Beckmann *et al.* (1983b). The concept is still useful today, however, some changes were added. As the center of the ellipse is crucial, different approaches were introduced. Whether taking the home location, the arithmetic mean of the geocoded data or merging the two main hubs within the data, the best centroid for the ellipses is not yet fully resolved (Schönfelder and Axhausen, 2004). A more detailed description about confidence ellipses is found in Section 3.2. In addition to the spatial research, the study of Schönfelder and Axhausen (2003) has investigated the differences between particular socio-demographic groups and whether elderly people experience social exclusions in any terms. However, it was found that neither gender, income nor age lead to significant differences.

2.2 Innovation Rate

Innovation is defined as new ideas, concepts and thoughts and as far as activities and locations are concerned, also new places. During a life time new locations are visited on a regular basis. Whether visiting a newly opened museum, traveling or eating out in a new restaurant around the corner. These newly seen places are added to the users repertoire of locations and therefore count towards innovation. As mentioned above, tracking locations is the crucial part in detecting new location. As far as research on innovation and innovation rates is concerned, not many studies have been performed, which is partially explained by the lack of available data. The Mobidrive data from 1999 and the later conducted study in Thurgau in 2003 (Löchl *et al.*, 2005) took a closer look into the innovation of the participants. As Schlich *et al.* (2004) showed, a strong regularity in travel behaviour over a prolonged observation period can be detected. This leads to the assumption that the number of places visited is unlimited and therefore the innovation rate never drops to zero, since the innovation behaviour stays constant. In other words, a growth in the number of locations can be observed with an increasing number of trips (Schönfelder and Axhausen, 2004). Figure 2 shows the observed innovation rates for several studies. However, research has shown that the rate of innovation drops with the

Figure 2: Comparing innovation rates across studies



Source: Schönfelder and Axhausen (2004)

length of observation and after some time, reaches a plateau. Furthermore, research shows that there are reasons to assume that the true innovation rate is lower than the values found. Due to the fact that certain activities have a low frequency, they might be detected as new, even though the participant has visited the place before, for example a dentists office (Schönfelder and Axhausen, 2004). In the Thurgau studies these frequencies were taken into account. In the post-processing the participants were asked if they had visited the places before. This additional filtering for new locations made it possible to have an innovation rate much closer to reality and lower than in other studies. The values given from the Thurgau data are around 0.3 ± 0.1 new locations per day whereas the Borlänge study found an average value of 0.5 new locations/day (Schönfelder and Axhausen, 2004). However, by examining the innovation rate some questions do still arise, which need to be further explored in future research. The locations need to be distinguished if they are totally new to the user or if they have visited the place before. Further the accuracy of the geodata must be improved in combination with clustering, so that a single activity is assigned to a corresponding cluster of geocode data rather than that locations are taken into account multiple times because of GPS drift or similar.

One solution to prevent this overcounting is by forming clusters, an aggregation of several location points and defining the respective activity to it. As simple as the concept might seem, the parameters used during this process may vary a great deal and estimating those is strongly dependent on the accuracy of the tracking. The cluster method by Anderberg (1973) is used in several studies and should not only bind the locations for the innovation but also for the activity space (Schlich *et al.*, 2004). One of the most recent studies regarding clustering and determine gatherings is Sekara *et al.* (2016). The aims of that paper is to predict travelers future behaviour based on its past and predicting the travelers next movements.

2.3 Overview of Previous Studies

This section will give an overview of what has been done so far. According to Axhausen and Schönfelder (2010) several studies have been conducted in order to get a better insight in the spatial human behaviour. Table 1 shows an overview of those studies and depict their characteristics.
Within this overview, it should be noted that the Borlänge study has by far the longest observation time, however the area for monitoring was limited and the sample of users is rather small. Furthermore, only private used cars were observed. Notwithstanding these limitations in the data acquisition process, the observations of the Borlänge study fundamentally influenced further research in this area.

Table 1: Summary of the values for the activity spaces

Name of the survey*	Year	Original focus	Locations(s)	Period	Resolution: geocoding	Resolution: purpses	Persons	Trips
Uppsala Household Travel Survey	1971	Travel behaviour	Uppsala, Sweden	35 days	Building	All purposes	144	23'000
Mobidrive: Dynamics and routines of travel behaviour	1999	Stability of temporal patterns	Karlsruhe and Halle, Germany	42 days	Street block	All purposes	361	52'000
Borlänge GPS study (ISA Rätt Fart)	2000-2002	Speeding behaviour	Borlänge, Sweden	Up tp 80 weeks	Trip ends: GPS; unique locations: pre-defined clusters of trip ends	Unknown, potentially all	189 veh**	240'000 car trips
Leisure study (SVI Gesetzmässigkeiten des Wochenend-Freizeitverkehrs)	2002	Leisure travel behaviour and activities	Zürich, Switzerland	84 days	Post-code level	31 leisure purposes	75	9'900 leisure activities
Thurgau diary (SVI Study of the stability of transport behaviour)	2003	Stability of temporal patterns	Frauenfeld and villages in the Swiss canton of Thurgau	42 days	Building	All purposes	230	37'000
Copenhagen GPS study (AKTA Road Pricing Experiment in Copenhagen)	2001-2003	Route choice under road pricing	Copenhagen, Denmark	18-24 weeks	Trip ends: GPS; unique locations: pre-defined clusters of trip ends	Unknown, potentially all	500 veh.	250'000 car trips
Atlanta GPS study (Commute Atlanta Study)	2004-2006	Travel behaviour; test of policy measures such as pricing	Atlanta, USA	Up to two years	Trip ends: GPS; unique locations: pre-defined clusters of trip ends	Unknown, potentially all	Approx. 500 veh.	Approx. 1'000'000 car trips

*In the following, the data sets are simply titled Mobidrive, Thurgau, Uppsala, Borlänge, Copenhagen and Atlanta for better readability.
**Private cars only

Source: Adopted from Axhausen and Schönfelder (2010)

As seen above, there have been attempts to determine and to quantify activity spaces and innovation rates, however the data acquisition process failed to fulfill the need for such comprehensive analyses. The project MOBIS (MOBIS, 2019) aimed to change hitherto lacking data quality issues. The unprecedented project set the foundation for a more sophisticated breakdown in order to investigate and compare individual user, as well as allowing for detailed analysis for various additional parameters such as socio-demographic groups.

3 Methods

This section is structured in three parts. The first part describes the Mobis dataset, which was investigated, the second part elaborates on the concept of activity spaces and how the 95 %-confidence ellipses are calculated. The last part describes the innovation rate and which approaches have been used for its computation. The calculations are done using R and RStudio. The code can be found in Appendix B.

3.1 MOBIS Data

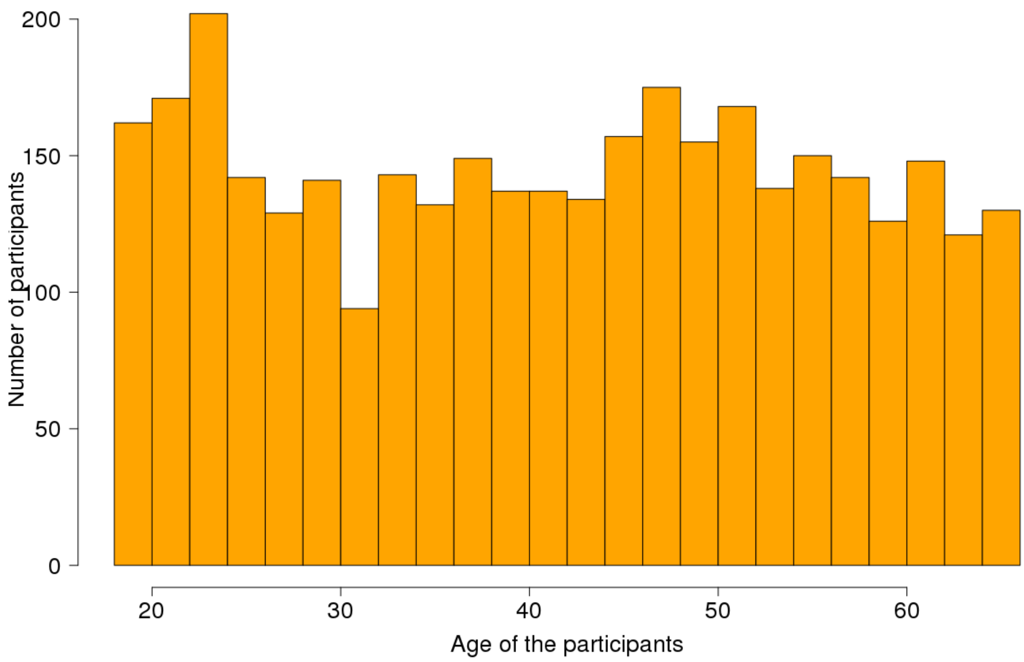
The data which is analysed originates from the research project «Mobility behaviour in Switzerland», or MOBIS. The project is a collaboration between the ETH Zurich, the University of Basel and the Zurich University of Applied Sciences, but funded and financially supported by the Swiss Innovation Agency (Innosuisse) and the Federal Department for the Environment, Transport, Energy and Communications (DETEC). The goal of this project is to investigate travel behaviour and thereby helping to improve the traffic structure and system in Switzerland. The project was initiated in 2018 and is still ongoing (MOBIS, 2019).

The data acquisition took place in several steps, consisting of surveys and tracking periods. The collections window is set to two months from the start of the participation. Having an ongoing study the participants were also asked to track their data during the COVID19 outbreack in Europe. However, this phase is going to be neglected here. MOBIS enrolled around 3'700 people which have already tracked their activities. The participants were grouped into three groups: the control, the nudging and the pricing group approximately a third of the sample each.

The tracking was carried out using a smartphone app, so that the geographical coordinates were transmitted as well. Those geodata was then extracted and builds the foundation of the calculations described in Section 3.2 and 3.3. For the analysis the different groups are computed separately as well as collectively. This gives the opportunity to reorder and to compare certain groups or aspects within the different groups. Furthermore, the activities have been filtered for activities taking place in Switzerland as well as for Phase 1 and 2 tracking. Since the requirements for the activity spaces and the innovation rates are different, there is a special group defined for the innovation rates. In order to see the variability in the innovation over the observation time span, the innovation for all the participants are computed. This also includes a determination of the number of observation days. As mentioned above, a longer observation period leads to a more accurate innovation rate. Therefore, the participants are filtered for only those, who have an observation time of at least 52 days. The time span of 52 days was chosen to have a more or less equally large sample size, roughly 1200, in the 'Innovation' as in the other groups. Those participants are saved in a separate file. Furthermore, some of the participants refused to answer certain questions. These were excluded for the corresponding analysis.

Having sorted and filtered for the different group several distributions were plotted. Whereas Figure 3 shows the distribution for the age over the whole set of participants, Figure 4 illustrates the distribution for the generated innovation group. The figures for the other groups regarding age distribution can be found in the Appendix A.1. It can be

Figure 3: Age distribution for all participants

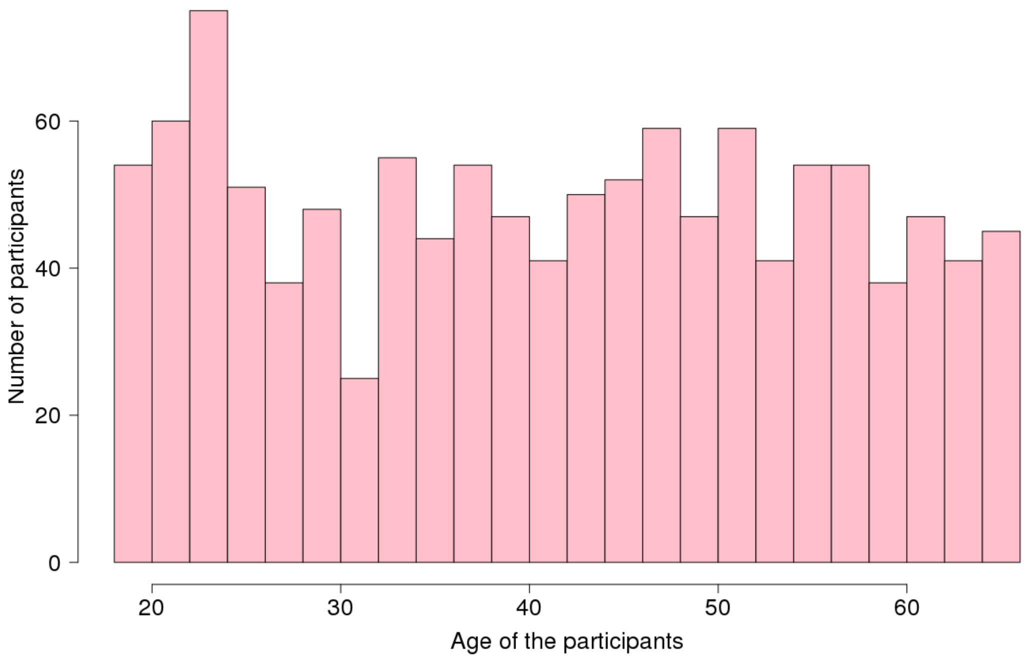


Source: MOBIS-Data

seen that the age distribution is rather uniform, except for the ages 22 to 24 and 30 to 32. However, a sufficient number of participants is achieved and therefore, the results are not to heavily biased to a particular age. The innovation group shows more significant discrepancies with the late 20s to early 30s being underrepresented.

The participants where then grouped by household size, which can be seen in Figure 5 for

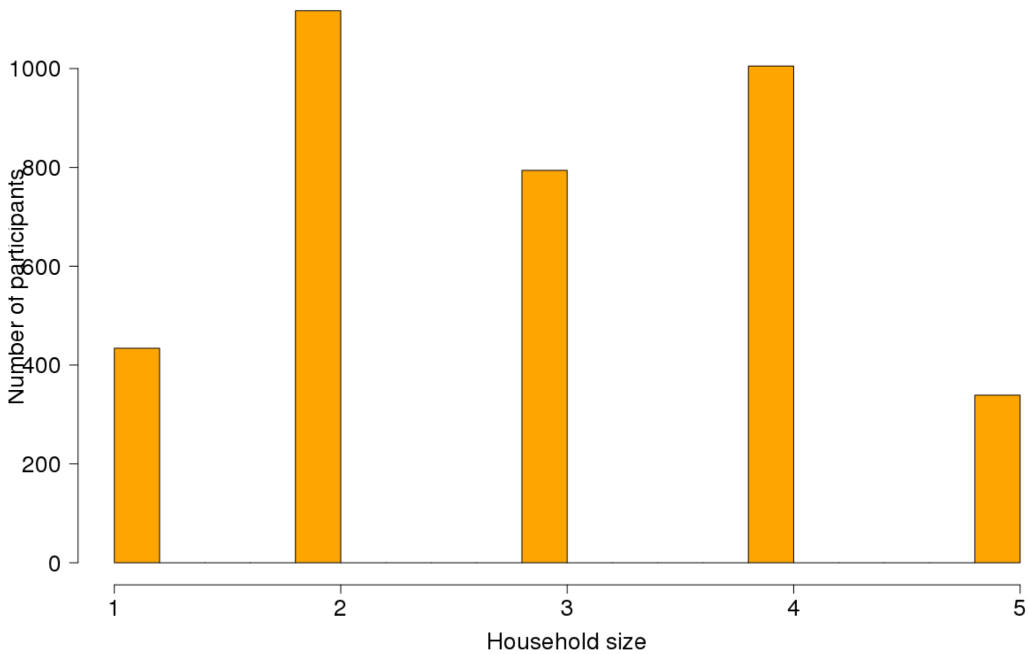
Figure 4: Age distribution of the innovation group



Source: MOBIS-Data

all participants. Due to the low variability in the groups, the subdivided plots are not included here, but can be found in the Appendix A.1.

Figure 5: Distribution of the household size over all participants



Source: MOBIS-Data

In the next classification, the income has been investigated. The participants were allotted to their income group, which stands for a certain amount of salary per month. The specifications can be found in Table 2 and Figure 6 depicts the distribution of those income groups for all participants. The subdivided groups can be found in the Appendix A.1.

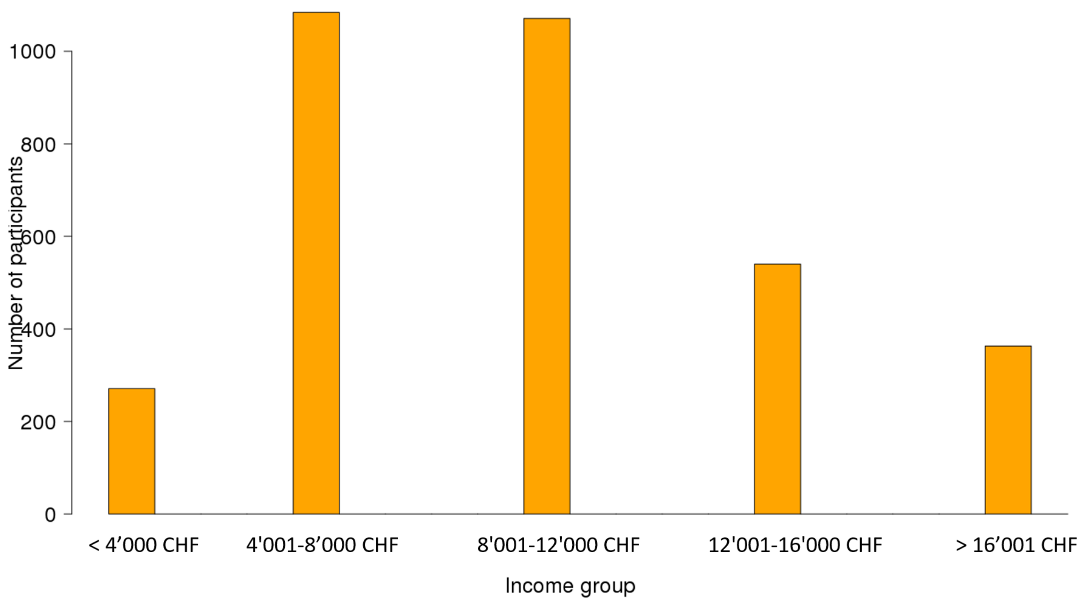
Table 2: Monthly salaries for income group 1 to 5 in CHF/month

Income group	1	2	3	4	5
Monthly salary	$<$ 4'000	4'001-8'000	8'001-12'000	12'001-16'000	$>$ 16'001

Source: MOBIS-Data

The distribution of participants regarding gender and language are depicted in Appendix A.1.

Figure 6: Distribution of the income over all participants



Source: MOBIS-Data

3.2 Calculation of Activity Spaces

This section is based on the literature and implementations of Schönfelder and Axhausen (2004). The corresponding R-code can be found in the Appendix B.2 and B.3.

An activity space is a spatial representation of the space in which a participant of the study moves. Taking all activity geodata for a single user, which fit the requirements of Section 3.1, a mean can be determined. This mean is equivalent to the arithmetic mean and can define the centroid of the ellipse. It should be noted that this centroid does not necessarily have to be a real point the user ever visited, but rather a conceptual point. To perform the analysis the activity coordinates need to be transformed to the Swiss reference system, projection code 2056. Following this transformation, the activity spaces can be computed. The method used is the 95%-confidence ellipse (Zahavi, 1979).
The centroid of the ellipse consists of the mean of all seen activity locations, weighted by

of duration for each corresponding activity. The weighting factor prevents a bias towards short activities. Furthermore, the covariance of the coordinates must be determined, in order to get the shape of the ellipses. These values can be calculated using the following formulas:

$$S = \begin{pmatrix} s_{xx} & s_{xy} \\ s_{yx} & s_{yy} \end{pmatrix} \tag{1}$$

These covariances s_{ij} are defined as followed:

$$s_{xx} = \frac{1}{n-2} \sum_{i=1}^{n} w_i * (x_i - \bar{x})^2$$

$$\mathrm{s}_{yy} = \tfrac{1}{n-2} \textstyle\sum_{i=1}^{n} w_i * (y_i - \bar{y})^2$$

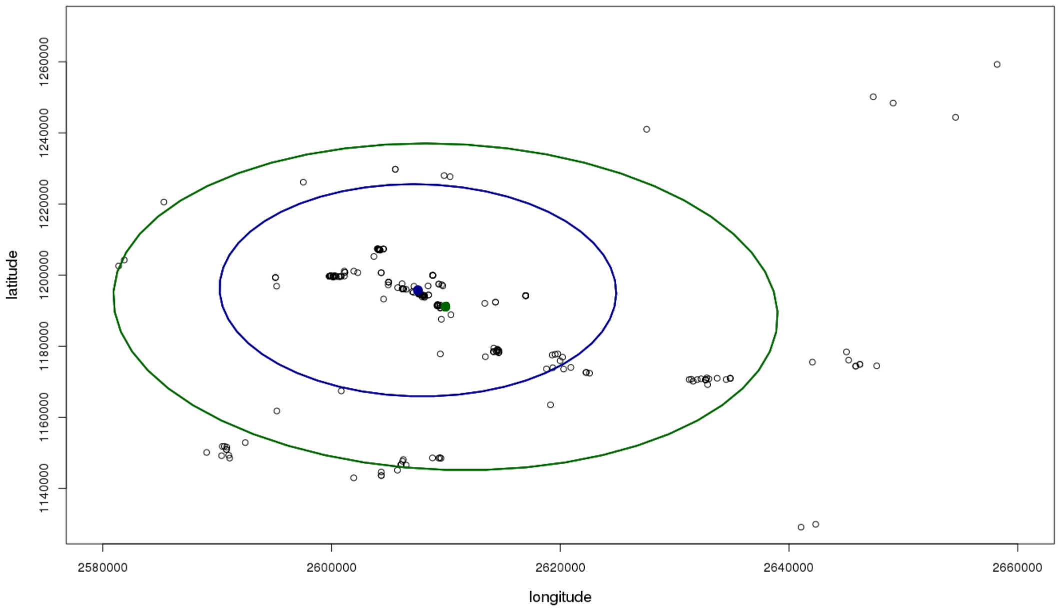
$$\mathrm{s}_{xy} = s_{yx} = \tfrac{1}{n-2} \textstyle\sum_{i=1}^{n} w_i * (x_i - \bar{x})(y_i - \bar{y})$$

w_i stands for the weight of the activity location, which is determined to be the activities' duration. Notice, if w_i is equal to 1 for all i, there is no weighting applied. Using those calculations the ellipses can be computed as well as plotted. Applying weights on the activity locations allocates the emphasis on the heavily used location. The longer a participants stays at a certain location the higher is its presumed importance. Figure 7 depicts the differences which arise by using weighted or unweighted locations. The green color represents the unweighted ellipse, whereas the blue indicates the weighted one. Moreover, the size of the ellipses can be determined by using Equation 2.

$$A = 6 * \pi \left|S\right|^{1/2}. \tag{2}$$

Those ellipses can graphically be represented, see Figure 7 and the figure displayed on the front page. The green ellipse depicts the unweighted activity space whereas blue shows the weighted.

Figure 7: Activity ellipses of one user. The green ellipse depicts the unweighted activity space whereas blue shows the weighted.



Source: MOBIS-Data

3.3 Calculation of Innovation Rates

The goal of the innovation rate is to determine how many new places are visited during a specific time frame. In order to apply a linear regression model on the data, each participant's innovation as well as its rate must be found. The computation is based on the approach of Schönfelder and Axhausen (2004). In a first step, a group of suitable participants needs to be found and chosen, as mentioned previously. This was done by selecting participants with an observation time span of at least 52 days. Therefore, each user's observation period has been determined. A long-term observation time frame allows to distinguish between the total innovation rate as well as the plateau rate, which settles in after the first few weeks.

In the first step of the computation, the activities are clustered. The method of clustering is further elaborated in the next paragraph. At this point it should be noted that activities and locations are equivalent. In other words, if a new location, not consisting of a cluster is detected, it is assumed that a new activity was performed. Forming these clusters and arranging them by the date it had been visited, it can be assigned to an ID which links the date of the first visit to the location. Having determined the clusters, the outliers can be identified. Due to the fact that each new activity added starts as an outlier and changes to a cluster over time, the outliers are assigned to an ID as well. Setting up a chronological list of IDs and its characteristics, it can be determined whether those IDs (locations) have been seen before or not. Finding those new activities is based on a day to day resolution in which two sequential days are compared. If a new location is detected, the program produces an output and sets the logical value to 'TRUE'. In order to compare the different innovation rates, the observation day must be standardized since all the participants have started at different dates throughout the year. The innovation rate can then easily be determined by calculating the mean of the data. From this point, the next step of the analysis can be approached. The distribution by age or income can be made and can be graphically depicted.

The clustering is separately explained here, due to the reason that it plays a crucial role in detecting the locations as such and not only random points. The clustering is performed independently from the computation of the innovations rates, however, the clustering function is integrated into the innovation program, see Appendix B.4. According to Schönfelder and Samaga (2003) cluster must be formed. In this analysis clusters are formed with the functions DBSCAN (Density-based spatial clustering of applications with noise) in R. The points which are defined are

- core points
- border points and
- outliers

The clusters are aggregated within a radius of 50 m, however, this can be changed to any given radius for which a cluster should be formed and the minimum number of points for a cluster is set to 3. This parameter was chosen to prevent any location to be automatically treated as a new cluster.

4 Results

Notice that not all results are depicted in this report. For further and more detailed insights referrer to the R-Code provided in the Appendix B and for a short explanation of the results, see Appendix C. Additionally it should be noted that in further depiction of the result not all outliers are shown in the figures in order to improve the readability of the plots. Whereas the activity spaces are given in the unit of [km^2], the innovation rate is in the unit of [New activities/day].

4.1 Activity Space

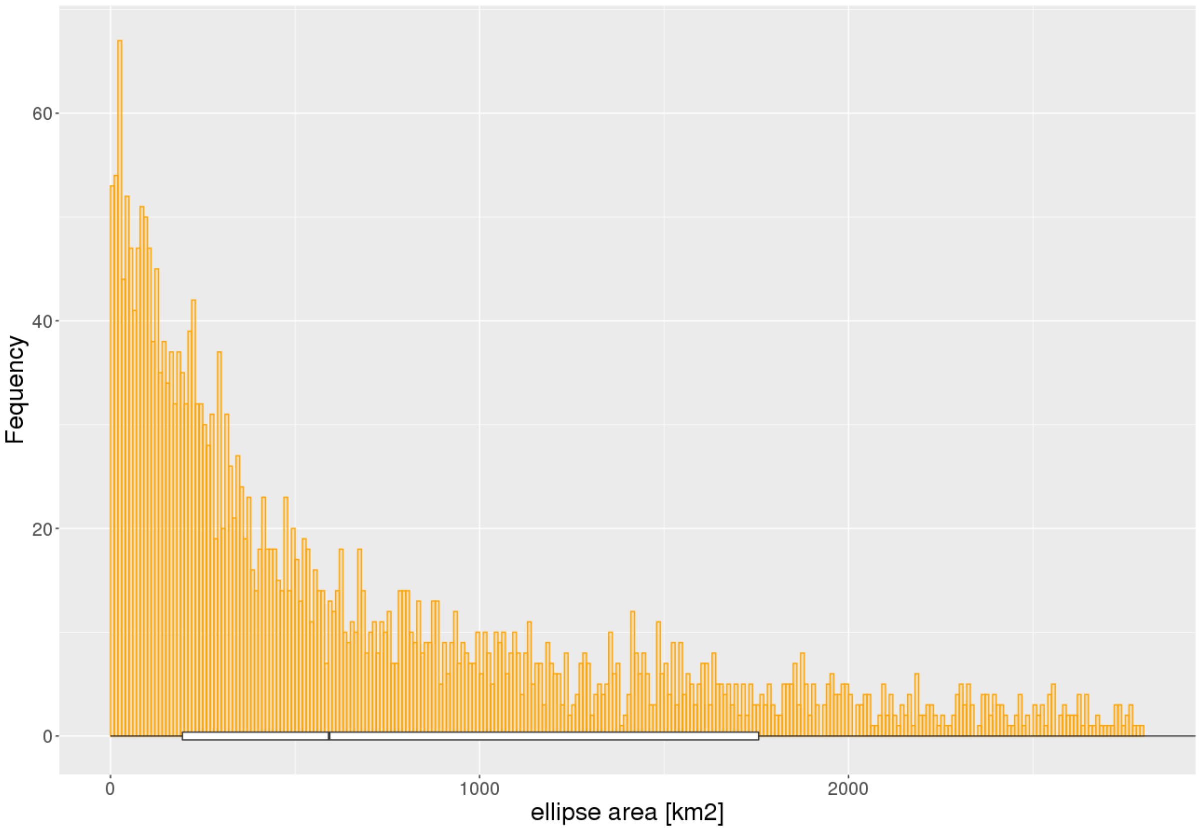
The results for all participants are shown in Figure 8 and Table 3. It can be seen that the standard deviation is relatively high, however, there is a tendency for smaller standard deviations when a larger sample size is taken into consideration. The standard deviation of all participants can be visualized by the size of the canton Vaud. Due to readability of the graphics, the whole data set is not shown. Nevertheless, the 25%-quantile lies roughly at 200 km^2 and 75%-quantile around 1800 km^2. As a comparison, these numbers approximately represent the lake of Neuchatel and the size of the Canton Zurich. It should be noted that for being included in the Innovation group a certain number of observation days must be reached, see Section 3.1.

Table 3: Summary of the values for the activity spaces for the different groups

	Control	Nudging	Pricing	Innovation	All Participants
Participants	1241	1245	1200	1241	3686
Mean	1628.8	1756.5	1873.5	1745.3	1754.2
Median	541.3	558.2	668.2	644.7	610.5
Std. deviation	± 2903.5	± 3531.6	± 3389.5	± 3272.2	± 3285.6
25%-quantile	189.8	194.0	203.1	215.3	189.8
75%-quantile	1705.1	1687.8	1880.2	1779.3	1705.1

Source: MOBIS-Data

Figure 8: Activity space distribution for all participants



Source: MOBIS-Data

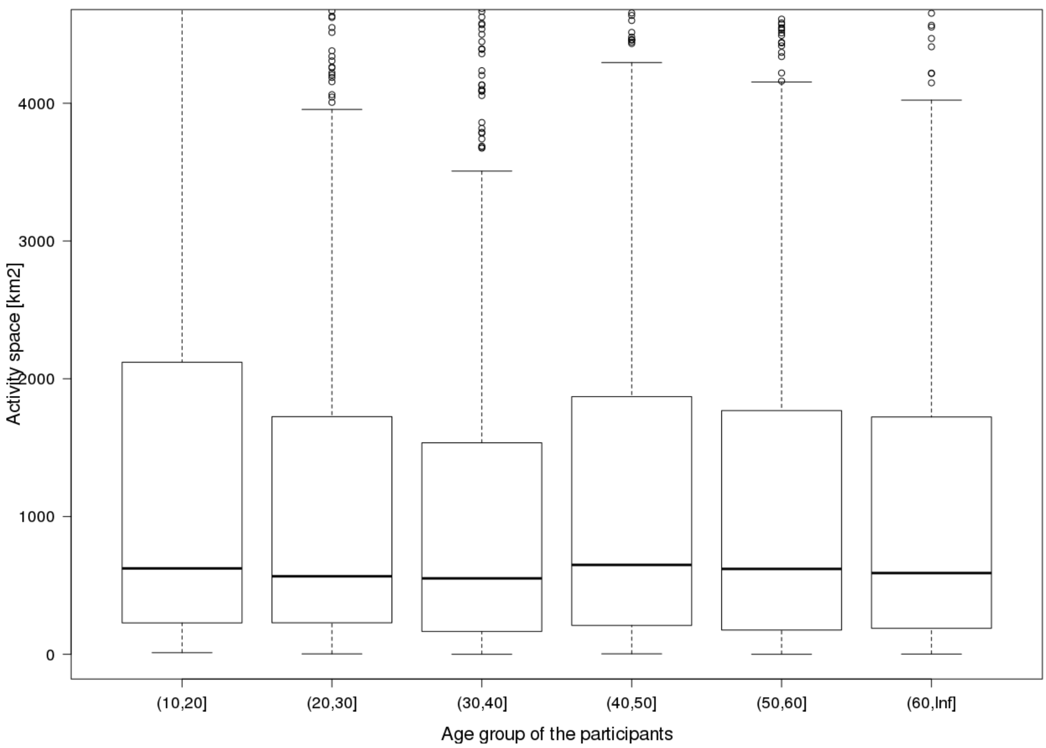
In order to get a better insight in the distribution of activity spaces, the MOBIS-Data has been explored regarding different aspects. Those are going to be shown below. In one the participants are grouped by **age**. The key figures such as the mean, median and the standard deviation can be found in Table 4 and 5. In these listings it can be seen that the standard deviation for lager samples is lower than for any of the treatment groups: control, nudging and pricing and also for the created innovation group. The graphical representation of the results in Figure 9 illustrate, that there is a variability in the different age groups and that especially the younger and the older participants tend to have a larger activity space area than the participants in their 20s and 30s. Moreover, it should be noted, that the number of participants for certain ages are relatively low and that the representation is more affected by one or two outliers. The more detailed analysis can be found in the Appendix A.1.

Table 4: Summary of the values for the activity spaces with known age by treatment group

	Control	Nudging	Pricing	Innovation	All Participants
Participants	1171	1174	1134	1179	3479
Mean	1574.1	1702.8	1897.7	1754.5	1737.8
Std. deviation mean	± 662.6	± 658.8	± 636.8	± 615.3	± 350.5
Median	557.8	561.1	675.1	626.6	563.9

Source: MOBIS-Data

Figure 9: Activity space distribution by age groups



Source: MOBIS-Data

Table 5: Key figures for activity spaces with known Age Groups

Age Groups	Number of participants	Mean	Median	Std. deviation
(10,20]	162	2042.8	623.3	± 3655.8
(20,30]	785	1670.8	566.2	± 2937.2
(30,40]	654	1630.0	551.1	± 3144.1
(40,50]	757	1795.3	649.0	± 3238.8
(50,60]	722	1707.8	619.7	± 3351.0
(60,Inf]	399	1908.2	589.6	± 3557.1

Source: MOBIS-Data

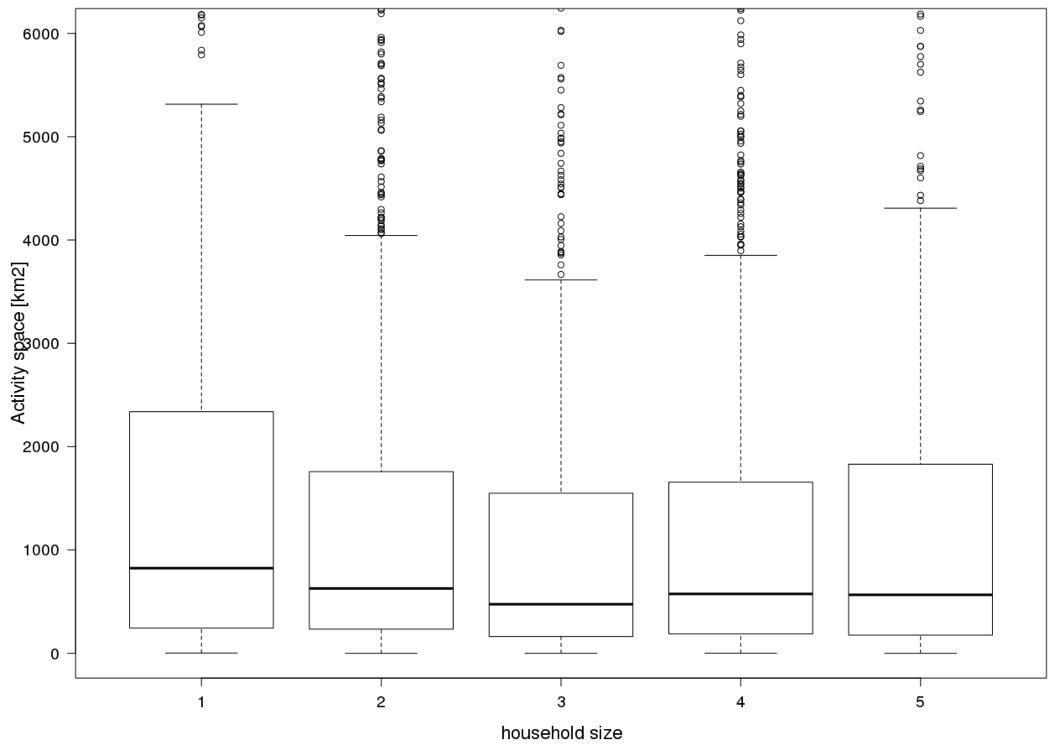
In the next graph the influence of **household size** is represented. Figure 10 shows the distribution of the size of activity spaces by household size. The smallest mean area can be detected at a household size of three. Furthermore, for a household size of three, the variability is the lowest. Additionally, the standard deviation decreases if the number of users increases. The key figures regarding the household size are listed in Table 6 and 7.

Table 6: Key figures for the activity spaces distribution with known household size

	Control	Nudging	Pricing	Innovation	All Participants
Participants	1241	1244	1200	1241	3685
Mean	1686.9	1776.5	1903.4	1700.4	1797.1
Std. deviation mean	± 275.4	± 514.0	± 183.0	± 223.0	± 218.8
Median	601.4	553.6	681.1	675.5	675.5

Source: MOBIS-Data

Figure 10: Activity space distribution by household size



Source: MOBIS-Data

Table 7: Key figures for activity spaces of different household sizes

Household Size	Number of participants	Mean	Median	Std. deviation
1	433	2162.2	823.9	± 3893.5
2	1117	1690.4	627.7	± 3212.3
3	793	1583.4	474.8	± 3129.5
4	1003	1766.0	574.5	± 3302.2
5	339	1783.7	565.9	± 2945.1

Source: MOBIS-Data

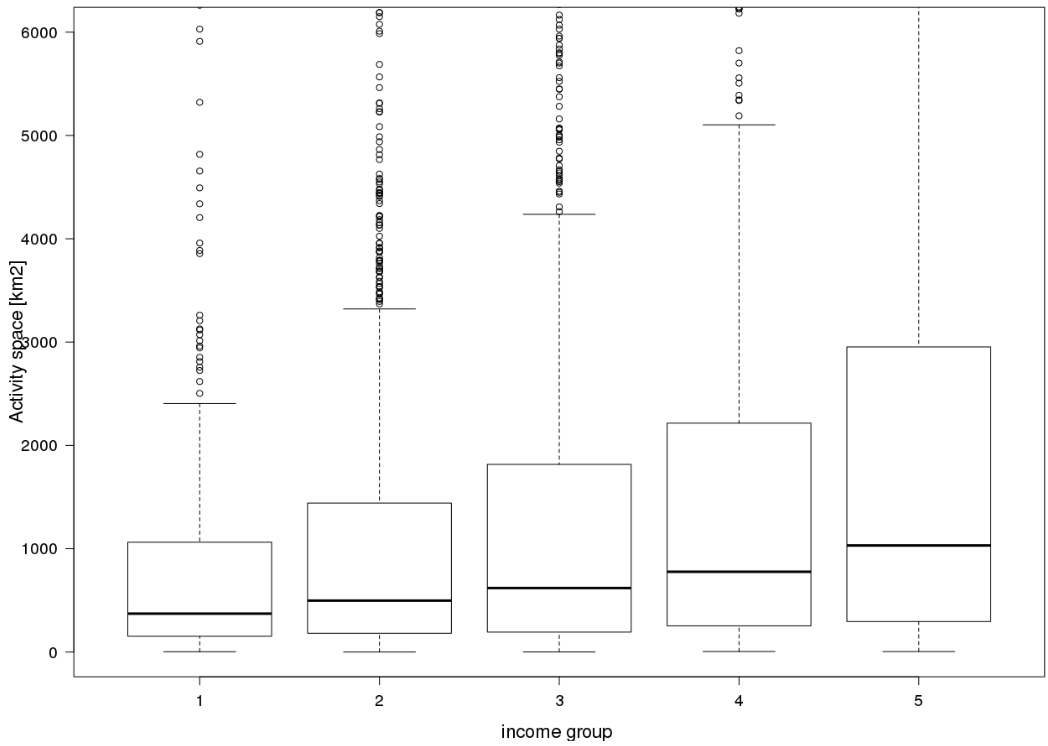
As far as **income** is concerned, the size of the activity space tend to grow by increasing income as depicted in Figure 11. The graphical representation shows that the variance within an income group is bigger with increasing income. Further, the key figures for the different income groups can be seen in Table 8 and 9.

Table 8: Summary of the values for the activity spaces grouped for income

	Control	Nudging	Pricing	Innovation	All Participants
Participants	1122	1131	1072	1125	3325
Mean	1718.1	1864.0	1925.3	1797.9	1832.5
Std. deviation mean	± 532.4	± 717.9	± 581.5	± 591.2	± 597.7
Median	573.9	612.5	656.6	668.1	638.9

Source: MOBIS-Data

Figure 11: Activity space distribution by income



Source: MOBIS-Data

Table 9: Key figures for activity spaces for different incomes

Income [CHF]	Number of participants	Mean	Median	Std. deviation
$<$ 4'000	271	1143.6	371.7	± 2290.2
4'001-8'000	1081	1402.7	497.6	± 2638.5
8'001-12'000	1071	1767.0	619.2	± 3303.2
12'001-16'000	539	2245.6	777.1	± 4028.0
$>$ 16'001	363	2603.6	1031.1	± 4213.1

Source: MOBIS-Data

To investigate if the "röstigraben" also influenced our mobility, the differences in travel behaviour regarding the **language** are conducted. The key figures in Table 10 and 11 show a trend that German speaking participants cover a larger average space than the French speaker do. Besides detecting a higher median for the German speaking group, also the standard deviation is higher, as shown in Figure 12.

Table 10: Summary of the values for the activity spaces by language

	Control	Nudging	Pricing	Innovation	All Participants
Participants	1241	1245	1200	1241	3686
Mean	1635.3	1801.1	1849.3	1727.1	1760.5
Std. deviation mean	± 18.3	± 132.7	± 86.1	± 78.4	± 27.3
Median	524.6	585.6	688.3	586.7	583.9

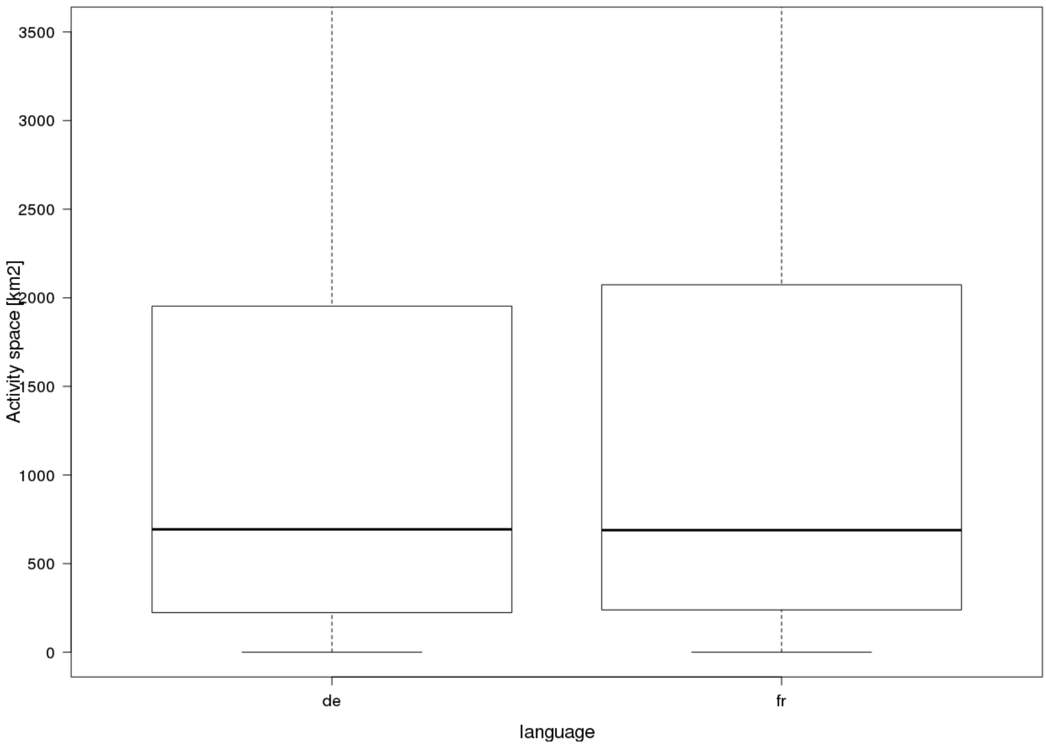
Source: MOBIS-Data

Table 11: Key figures for activity spaces for the different languages

Language	Number of participants	Mean	Median	Std. deviation
German	2688	1741.174	592.1001	± 3309.306
French	998	1779.784	575.6901	± 3222.464

Source: MOBIS-Data

Figure 12: Activity space distribution by language

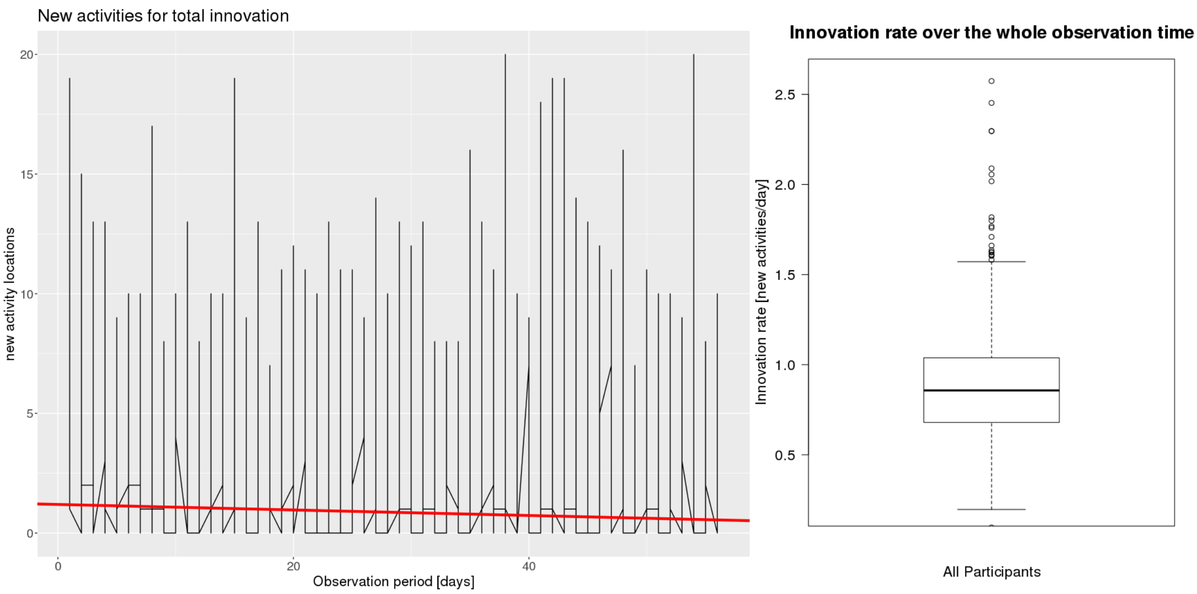


Source: MOBIS-Data

4.2 Innovation Rate

The innovation results are produced from two different points of view. The innovation rate has been computed for the whole observation time frame and for a more detailed analysis for the time span after 28 days of observation. An overview over the total observation time is given in Figure 13 and in Table 12, in which can be seen that for the overall innovation rate, the values are higher than for the second, later part of observation period. Those values should represent the plateau phase, which was mentioned before. It should be noted that the standard deviation over the total participation window is lower compared to the one only considering the days after the initial 28 days period. The difference is traced back to unequal days of observation for each user. The higher the day of observation the fewer participants have still been tracking. Whereas the average of the long-term observation is around 0.875 new activities per day, the mean for the second part of the observation period is 0.756 new activities per day.

Figure 13: Innovation rate over the whole observation period



(a) New activities (b) Rate distribution

Source: Mobis Data

Table 12: Innovation rate

	Total observation time	After 4 weeks of observation
Sample size [Participants]	1241	1241
Mean	0.875	0.756
Median	0.855	0.720
Std. deviation	± 0.289	± 0.302
25%-Quantile	0.685	0.560
75%-Quantile	1.04	0.920

Source: MOBIS-Data

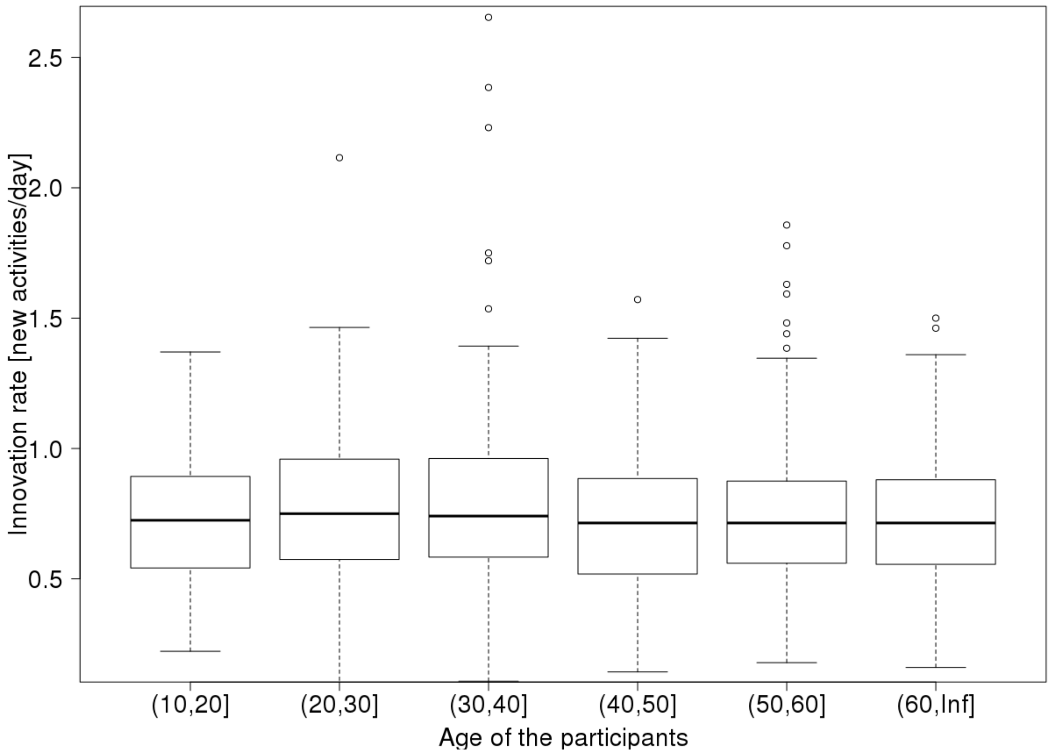
In order to get a constant representation of the innovation, in which the values are less biased on the starting period of observation, for further analysis only the innovation rate after four weeks are taken into account. The values in Table 13 show that the variability between the different **age** groups is relatively small. However, it should be noted that the middle aged participants have by far the highest innovation rate. Not only do they have the highest innovation, but also the highest standard deviation. As Figure 14 shows, the interquartile range are comparable in size for all age groups.

Table 13: Innovation rate after 4 weeks of observation for different age groups

Age Group	Number of participants	Mean	Median	Std. deviation
(10,20]	54	0.733	0.725	± 0.265
(20,30]	272	0.763	0.75	± 0.276
(30,40]	225	0.804	0.741	± 0.376
(40,50]	249	0.724	0.714	± 0.259
(50,60]	246	0.744	0.714	± 0.276
(60,Inf]	133	0.741	0.714	± 0.312

Source: MOBIS-Data

Figure 14: Innovation rate by age



Source: MOBIS-Data

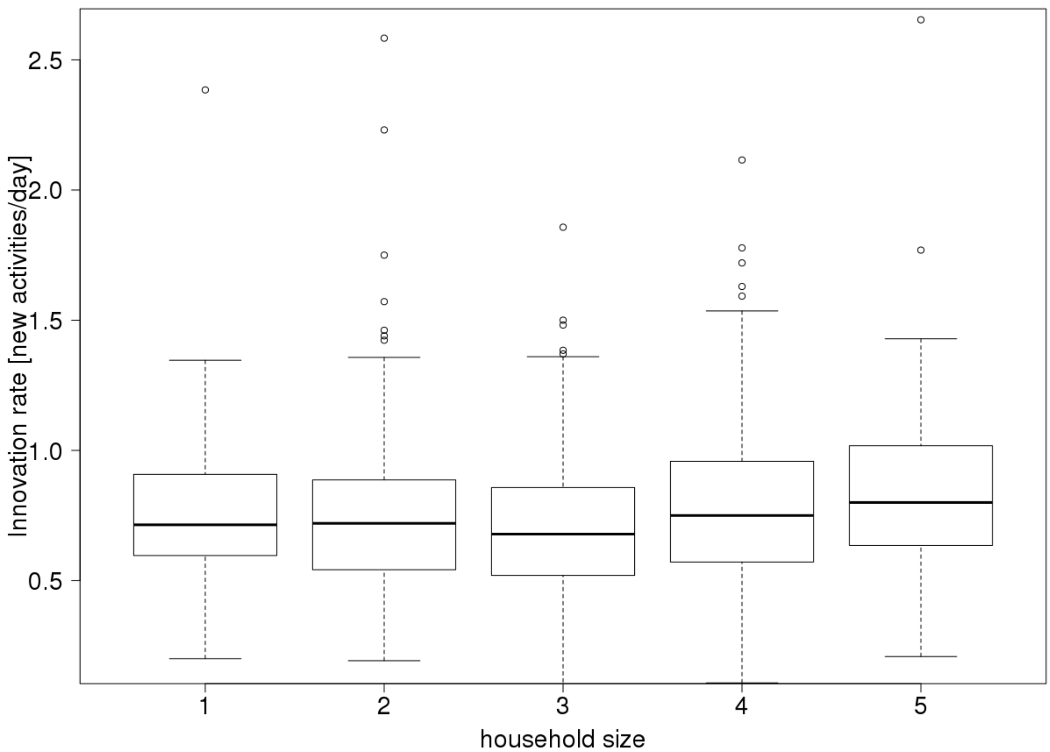
In terms of the **household size**, a minimum innovation rate can be found for three inhabitants within one household. Further, there is a tendency that, a higher number of people supports a higher innovation rate. As the key figures in Table 14 indicate, the variability within the groups increase by a growing household size. Figure 15 shows a graphical depiction of the results found regarding the dependence on the household size.

Table 14: Innovation rate after 4 weeks of observation regarding the household size

Household size	Number of participants	Mean	Median	Std. deviation
1	131	0.758	0.714	± 0.273
2	344	0.747	0.72	± 0.319
3	266	0.704	0.679	± 0.263
4	373	0.780	0.75	± 0.311
5	127	0.821	0.8	± 0.318

Source: MOBIS-Data

Figure 15: Innovation rate by household size



Source: MOBIS-Data

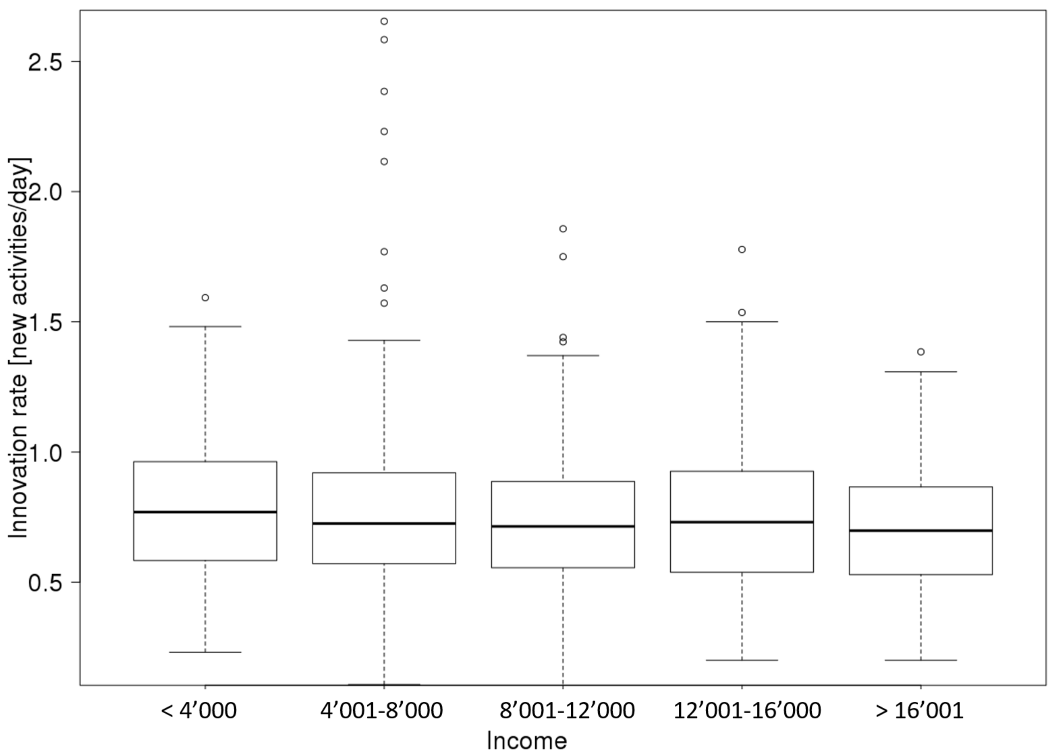
As far as the **income** is concerned, the variability is small, as can be seen by the standard deviation in Table 15. Whereas the interquartile rage of the innovation rate stays almost constant for all income group, the mean and the median tend to decrease with a higher income. The standard deviation indicates that the distribution within the groups are similar, however, there are still outliers which have an overproportional high innovation rate, as some of them can be depicted in Figure 16.

Table 15: Innovation rate after 4 weeks of observation regarding income

Income [CHF]	Number of participants	Mean	Median	Std. deviation
$<$ 4'000	106	0.777	0.769	$\pm$ 0.266
4'001-8'000	378	0.767	0.725	$\pm$ 0.343
8'001-12'000	356	0.748	0.714	$\pm$ 0.296
12'001-16'000	181	0.751	0.731	$\pm$ 0.273
$>$ 16'001	104	0.716	0.698	$\pm$ 0.258

Source: MOBIS-Data

Figure 16: Innovation rate by income



Source: MOBIS-Data

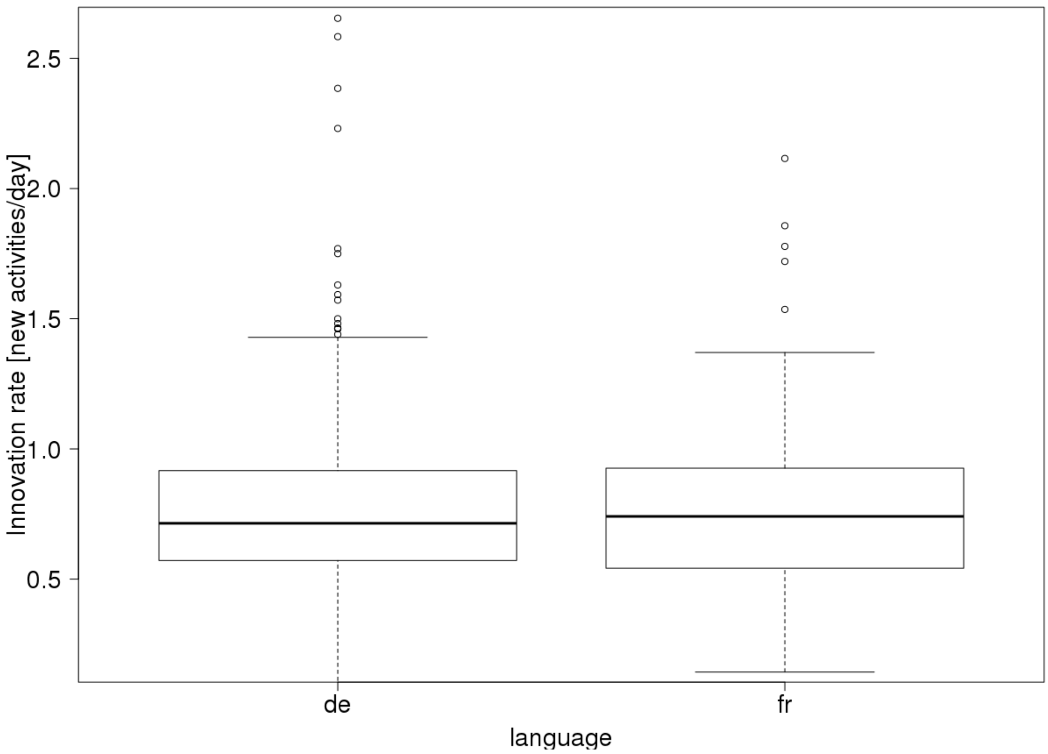
When it comes to the innovation rate regarding the **language** of the participants, there is a tendency that the French do innovate slightly more than their German counterparts. Figure 17 and Table 16 show the findings in that subcategory. Whereas the interquartile range for both, the French and the German speakers are similar, the mean and the median show higher values for the French share.

Table 16: Innovation rate after 4 weeks of observation

Language	Number of participants	Mean	Median	Std. deviation
German	923	0.753	0.714	± 0.292
French	318	0.766	0.741	± 0.331

Source: MOBIS-Data

Figure 17: Innovation rate by language



Source: MOBIS-Data

5 Discussion and Conclusion

5.1 Activity Space

Taking a look at the values found for the size of the activity spaces for all participants, it can be seen that the standard deviation is rather high. However, this is not surprising due to the fact, that the participants are humans and behave all in different manners. In a country such as Switzerland where commuting plays an important role, it is expected that certain attendee of the study show extraordinarily large activity spaces. Moreover, the standard deviation only slightly decreases for a bigger sample size. Reasons are as above, where some participants' environment is very limited whereas others cross the whole country multiple times. Further, it is expected, that the standard deviation decreases if more similar groups are found. Looking at the values of the interquartile ranges for the various groups, it can be seen that these are almost consistent for the subsets and for all participants.

As far as the age of the participants is concerned, there are some trends, which can be seen. Especially for the participants being in their 30s. It can easily be seen that over all subset of participants, the age group from 30 to 40 has the smallest activity space. Not only is the median the smallest from all of the groups but also the variability within the subset. Further, the participants group in their 30s represent the smallest interquartile range. Having said that, the observed results are contrary to what has previously been expected. The preconception was that the activity space of middle aged participants are the highest. Therefore, this finding is rather astonishing. This raises the question, why this value is comparatively low. However, there are several possible explanations, such as the low number of participants for that age group or that children are involved. It cannot be fully concluded what the main reason was, however, in order to quantify the assumption that children might reduce the activity space of parents the distribution of activity spaces regarding the household size was investigated next.

Considering the classification by the household sizes, Figure 10 shows that the highest activity space is covered by the single and double person households. This is as expected, since it is assumed that three-person households include at least one child. However, the MOBIS-Data does not answer the question how the household are composed, therefore it cannot universally be concluded if children reduce the parental activity space. It does make sense, that their space condenses, since the most frequented facilities used by children are schools, kindergartens, friends places or playing yards within a small radius from the domestic location. It would further explain why the activity space for the participants in the 40s starts to increase again. As children get older and become more and more independent, parents gain the opportunity to travel for their own purposes more often. In order to verify the influence of children, more research must be done. Moreover, dependencies between a child and parents should be explored by comparing children's travel dairies to that of their guardians. It then can be figured out how the behaviour changes in respect to the age of the child. However, this requires a comprehensive and overall family study over the time span of several years time. Even with today's technologies, this would be a tremendous effort - and have significant privacy implications.

Another aspect which was investigated was the effect of the income on the size of the activity space. The findings are in line with expectations. Over all the subset groups a tendency can be detected. Increasing income leads to a larger activity space. The assumption is that having a higher monthly salary means that travel cost for commuting to several locations does not play a restrictive role. Moreover, some leisure activities may be performed, which are simply not possible in the close environment because special requirements need to be met for activities such as golf, paragliding, flying etc. It is assumed that those detected trends are driven by a multitude of reasons. However, in order to conduct a more sophisticated study on why high income group participants travel within a larger area the purposes of trips must be known. What drives or even forces the participants to enlarge their space. The research field of trip purpose should be tackled so that not only this question could be answered, but also to determine the chain of effect. Do they travel because they have money or do they have the money because they travel?

The values for the subset of language distribution surprises. Comparing the standard deviation, the low values stand out. However, taking a more detailed look at it, the number of participants for each group are much larger than for the specific age group, for example. The ratio between German and French speaking participants is 923:318. Therefore it follows logically that the interquartile range tends to be larger for the German speakers. As the standard deviation should decreasing with increasing number of participants. There are no other observable tendencies. In terms of being able to detect any abnormality within a language group, in case there is any, it must be a huge discrepancy over the whole data set and something that affects everyone. At this point of investigation, this seems not to be the case in the MOBIS-Data.

5.2 Innovation Rate

The innovation rate presents itself very much as expected. As know from previous research, the innovation rate drops with time and reaches a plateau, where the rate is nearly constant onward. Both of these values could be found, however compared to previous conducted research, e.g Borlänge or the Thurgau study, the values are rather high. Despite the extensive data acquisition of MOBIS, some crucial details are missing. Especially when the actual innovation rate should be determined, it should be distinguished if a location already has been visited before the observation started. Since these information is missing, it can be assumed that the calculated innovation rates are a great deal higher than the real values. In post-processing of the Thurgau study, the participants were asked if the locations they visited were truly new locations, or if they have already been there before. Notwithstanding the missing data, the results which have been found prove the concepts from earlier work. Considering the second half of the observation time span, it can be shown that the rate changes over time and that in daily life a nearly constant rate is reached. However, Figure 13 shows a slight increase of new activity location in the second part of the observation time frame. This can have different reasons such as a sudden change in weather, changing of home location or traveling to name just a few. Overall, these increases in new activities have in common that the participant's desire to explore the environment around its comfort zone rises.
Comparing the values from the total observation time and the one starting in week 5, it should be noted, that the absolute numbers of the interquartile ranges are almost the same. Which means that the participants continuously behave similar within the observation time frame.

Comparing the activity space to the innovation rate regarding the separation of age, the results show a dependency of size and innovation. Whereas the activity space is the smallest for participants in their 30s, the innovation rate is the highest. This might be traced back on the limitations a single user has. The participants might be bound to a certain place because of family or work and therefore innovate more in order to have the same satisfaction as others. Another aspect might be that especially young people move in groups. Due to education and other activities they are in touch with many different people from different locations.
Additionally, it can be explained that young adults have not had the chance to see as many places as participants who are two generations older. It is assumed that the older generation has already settled, whereas the young adults want to travel and keep exploring. For further research it would be interesting to determine the reasons why the participants travel and how their behaviour is influenced by the integration of their respective society

and environment. As far as the MOBIS-Data is concerned, these information is missing. For future research it might be interesting to know the motivation why the participants explore new places and activities and why they start enlarge their comfort zone.

Furthermore, for participants living in a household with three people, a correlation for activity space and innovation has been discovered. Activity space as well as the innovation rate tend to be small respectively low. As mentioned above, the small activity space corresponding to a household size of three might indicate that there is at least one child living in that particular household. Having children changes the structure of the parents' life. Whereas at the beginning many new places were added to their dairies, after some time the daily life gets repetitive and mostly consists of a child dependent pattern. This could explain why both of this key figures decrease for the subset of a three inhabitant household. In order to have a better insight into the topic and determine the influence of children, corresponding data must be acquired.

Another interesting finding is detected in the group of low income. It is found that the smallest activity space is assigned to the highest innovation rate. It is understandable that if the income is low, the user might not be able to afford high travel expenses and therefore is forced to stay within a certain area. However, a participant might also be bound to certain location because of work or education. The correlation between small activity space and high innovation rate for a low income group should further be investigated in terms of the participants background. In the scope of this work, different work loads and how time is allocated to specific activities was not investigated. Low income does not necessarily mean that a job's salary is low, it can also imply that only part-time work is performed. In further research with the MOBIS-Data, this topic can easily be approached and should definitively been explored in the future.

The results on the innovation rate for different speaking participants does not show a trend nor any conspicuousness as previously thought since it could easily be assumed that a different kind of lifestyle and culture would comes with another spoken language. It was assumed that some tendencies would be observable, however, as it seems the French speaking part is not that different from the German speaking part as far as travel behaviour is concerned.

5.3 Conclusion

In conclusion, there are different approaches, which can be tackled in future research. Crucial questions have arisen, such as if children reduce the activity space and innovation rate of their parents. Moreover, the innovation results should be analysed more profoundly in terms of the actual innovation rate and the observed rate. Additionally, the influence of accessibility may play an important role in how innovation rates develop and change over time.

Concerning human behaviour and spatial distribution of individuals life, research has just started and the limitation for studies seem endless. The aspects of daily life and how these aspects influence the decisions made are further to be explored and discovered. Conducting more research in this area is warranted to predict mobility patterns and movements as precisely as possible, as such knowledge could play a crucial role in future planning of public infrastructure.

6 Acknowledgements

I would like to express my gratitude to my supervisor Prof. Dr. Kay W. Axhausen who gave me the possibility to pursue this semester project. His continuous inputs and encouragement were of utmost value for the project. Furthermore, I would like to express my thanks to Joseph Molloy for his tremendous efforts in supporting me throughout this work. His untiring support and feedback on the coding were crucial for the success of this project.

Last but not least, I would also like to thank my family and friends for their unconditional support during my studies and the time of this project. I specifically would like to thank Manuel Rytz for his feedback on the manuscript.

7 References

Anderberg, M. R. (1973) *Cluster Analysis for Applications - 1st Edition*, Academic Press.

Axhausen, K. (2002) A dynamic understanding of travel demand: A sketch, *Arbeitsberichte Verkehrs- und Raumplanung*, **vol 119**, May 2002.

Axhausen, K. W. and S. Schönfelder (2010) *Urban Rhythms and Travel Behavior*, Farnham Ashgate Publishing Ltd.

Axhausen, K. W., A. Zimmermann, S. Schönfelder, G. Rindsfüser and T. Haupt (2002) Observing the rhythms of daily life: A six-week travel diary, *Transportation*, **vol 29**, pages: 95–124, May 2002.

Beckmann, M. J., T. F. Golob and Y. Zahavi (1983a) Travel probability fields and urban spatial structure: 1. theory, *Environment and Planning A: Economy and Space*, **vol 15** (5) pages: 593–606.

Beckmann, M. J., T. F. Golob and Y. Zahavi (1983b) Travel probability fields and urban spatial structure: 2. empirical tests, *Environment and Planning A: Economy and Space*, **vol 15** (6) pages: 727–738.

Biding, T. and G. Lind (2002) Intelligent speed adaptation (isa) results of larger-scale trials in borlange, lidkoping, lund and umea during 1999-2002, *Statens Vaegverk /Sweden*, **vol 89 E**.

Brown, L. A. and E. G. Moore (1970) The Intra-Urban Migration Process: A Perspective, *Geografiska Annaler. Series B, Human Geography*, **vol 52**, pages: 1–13.

Dijst, M. (1999) Two-earner families and their action spaces: A case study of two dutch communities, *GeoJournal*, **vol 48**, pages: 195–206.

Downs, R. M. and D. Stea (1977) *Maps in Minds: Reflections on Cognitive Mapping*, Harper & Row.

Dürr, H. (1979) *Planungsbezogene Aktionsraumforschung : theoretische Aspekte und eine empirische Pilotstudie*, Beiträge / Akademie für Raumforschung und Landesplanung, Schroedel, Hannover.

Golledge, R. G. (1999) *Wayfinding Behavior: Cognitive Mapping and Other Spatial Processes*, JHU Press.

Golledge, R. G. and R. J. Stimson (1997) *Spatial Behavior: A Geographic Perspective*, Guilford Press, January 1997.

Horton, F. E. and D. R. Reynolds (1971) Effects of Urban Spatial Structure on Individual Behavior, *Economic Geography*, **vol. 47**, pages 36–48.

Lenntorp, B. (1976) *Paths in Space-time Environments: A Time-geographic Study of Movement Possibilities of Individuals*, Royal University of Lund, Department of Geography.

Lynch, K. (1960) *The Image of the City*, The MIT Press.

Löchl, M., S. Schönfelder, R. Schlich, T. Buhl, P. Widmer and K. W. Axhausen (2005) Untersuchung der Stabilität des Verkehrsverhaltens: Schlussbericht, *Technical Report*, ETH Zurich.

Maier, J., R. Paesler, K. Ruppert and F. Schaffer (1977) *Sozialgeographie*, Westermann, Braunschweig.

MOBIS (2019) MOBIS - Mobility Behaviour in Switzerland - Research Project, `https://ivtmobis.ethz.ch/mobis/en/`. Last access: 2020-05-26.

Schlich, R., S. Schönfelder, S. Hanson and K. W. Axhausen (2004) Structures of Leisure Travel: Temporal and Spatial Variability, *Transport Reviews*, **vol 24** (2) pages: 219–237, March 2004. Publisher: Routledge.

Schönfelder, S., K. Axhausen, N. Antille and M. Bierlaire (2002) Exploring the potentials of automatically collected GPS data for travel behaviour analysis - a Swedish data source, *GI-Technologien für Verkehr und Logistik*, (13) pages: 155–179. Publisher: Institut für Geoinformatik, Universität Münster.

Schönfelder, S. and K. W. Axhausen (2003) Activity spaces: measures of social exclusion?, *Transport Policy*, **vol 10** (4) pages: 273–286, October 2003.

Schönfelder, S. and K. W. Axhausen (2004) Structure and innovation of human activity spaces, *Arbeitsberichte Verkehrs- und Raumplanung*, **258**, p. 41.

Schönfelder, S. and U. Samaga (2003) Where do you want to go today?: More observations on daily mobility, paper presented at the *STRC 2005 conference proceedings : 5th Swiss transport research conference : Monte Verità/Ascona, March 9-11, 2005*, vol. 179.

Sekara, V., A. Stopczynski and S. Lehmann (2016) Fundamental structures of dynamic social networks, *Proceedings of the National Academy of Sciences*, **vol 113** (36) pages: 9977–9982.

Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis*, CRC Press, April 1986.

Zahavi, Y. (1979) *The "UMOT" Project*, US Department of Transportation, Washington.

8 Eigenständigkeiterklärung

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Eigenständigkeitserklärung

Die unterzeichnete Eigenständigkeitserklärung ist Bestandteil jeder während des Studiums verfassten Semester-, Bachelor- und Master-Arbeit oder anderen Abschlussarbeit (auch der jeweils elektronischen Version).

Die Dozentinnen und Dozenten können auch für andere bei ihnen verfasste schriftliche Arbeiten eine Eigenständigkeitserklärung verlangen.

Ich bestätige, die vorliegende Arbeit selbständig und in eigenen Worten verfasst zu haben. Davon ausgenommen sind sprachliche und inhaltliche Korrekturvorschläge durch die Betreuer und Betreuerinnen der Arbeit.

Titel der Arbeit (in Druckschrift):

Activity Spaces and Behavioral Innovation

Verfasst von (in Druckschrift):
Bei Gruppenarbeiten sind die Namen aller Verfasserinnen und Verfasser erforderlich.

Name(n):	**Vorname(n):**
Rytz	Sara

Ich bestätige mit meiner Unterschrift:
- Ich habe keine im Merkblatt „Zitier-Knigge“ beschriebene Form des Plagiats begangen.
- Ich habe alle Methoden, Daten und Arbeitsabläufe wahrheitsgetreu dokumentiert.
- Ich habe keine Daten manipuliert.
- Ich habe alle Personen erwähnt, welche die Arbeit wesentlich unterstützt haben.

Ich nehme zur Kenntnis, dass die Arbeit mit elektronischen Hilfsmitteln auf Plagiate überprüft werden kann.

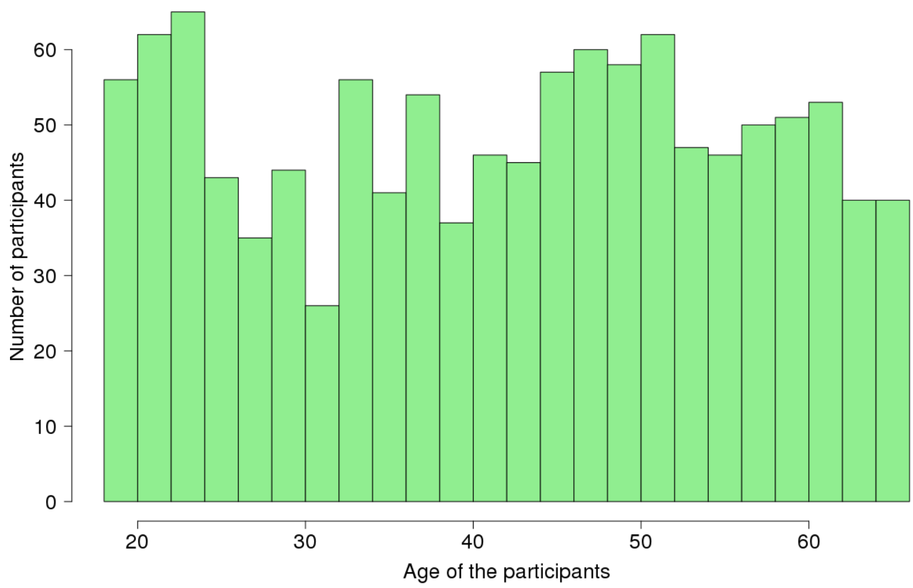
Ort, Datum	**Unterschrift(en)**
Rombach, 30.5.2020	S. Rytz

Bei Gruppenarbeiten sind die Namen aller Verfasserinnen und Verfasser erforderlich. Durch die Unterschriften bürgen sie gemeinsam für den gesamten Inhalt dieser schriftlichen Arbeit.

A Plots

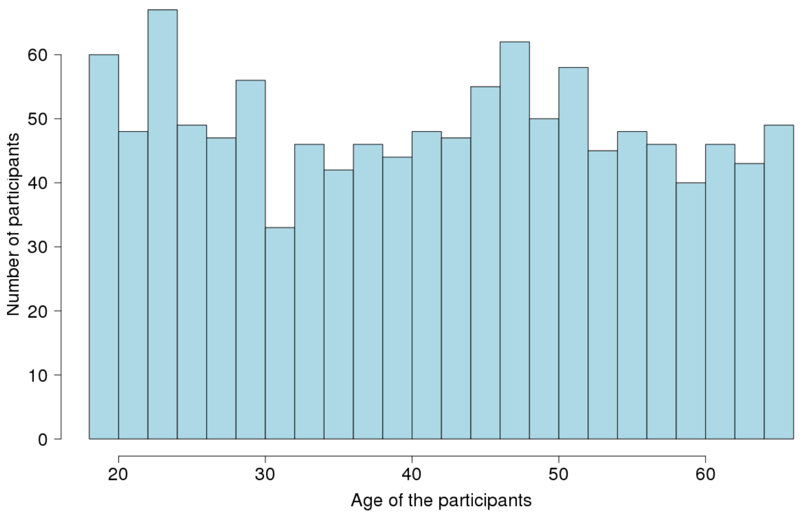
A.1 MOBIS Plots - RStudio

Figure 18: Age Distribution of the control group



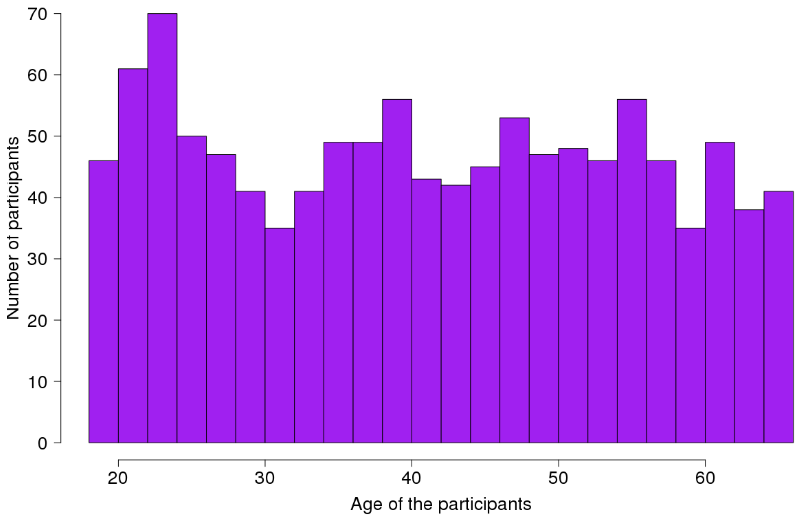
Source: MOBIS-Data

Figure 19: Age Distribution of the nudging group



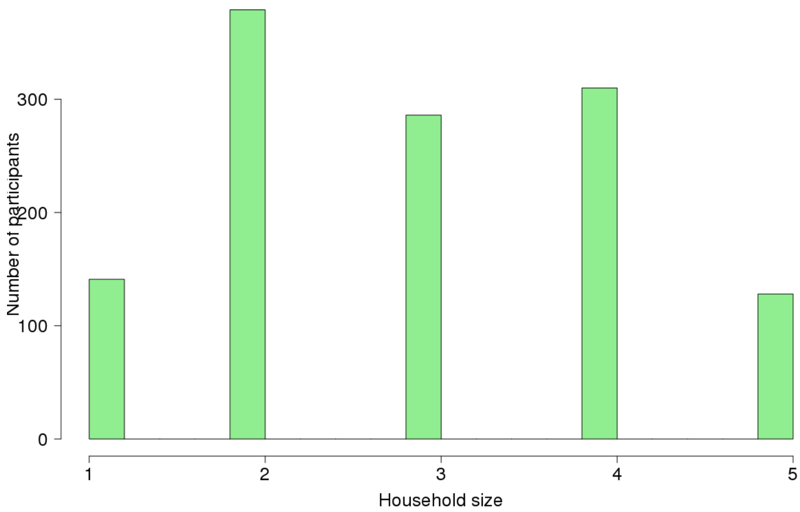
Source: MOBIS-Data

Figure 20: Age Distribution of the pricing group



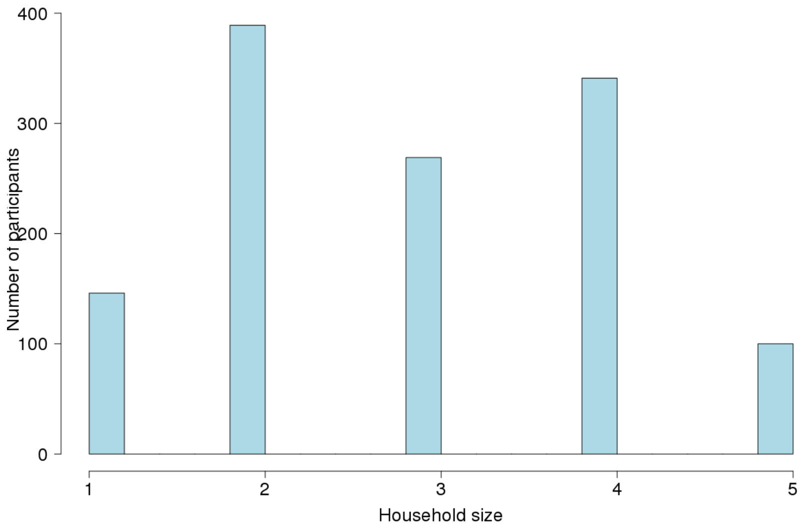
Source: MOBIS-Data

Figure 21: Household size Distribution of the control group



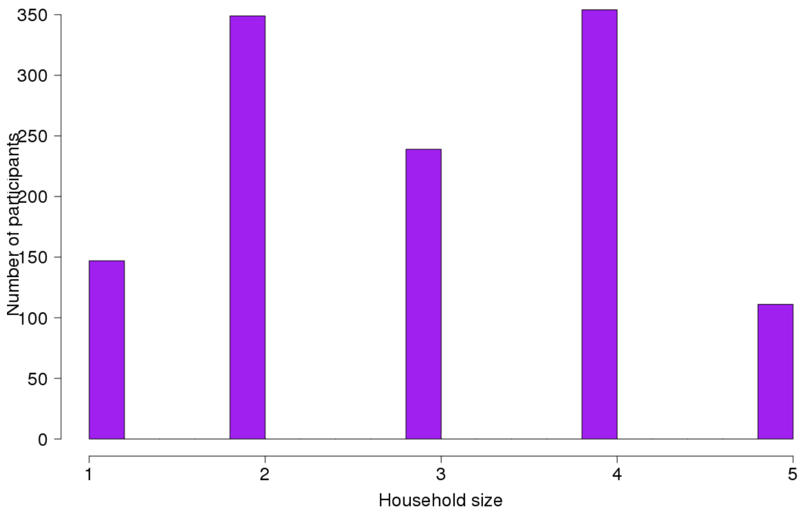
Source: MOBIS-Data

Figure 22: Household size Distribution of the nudging group



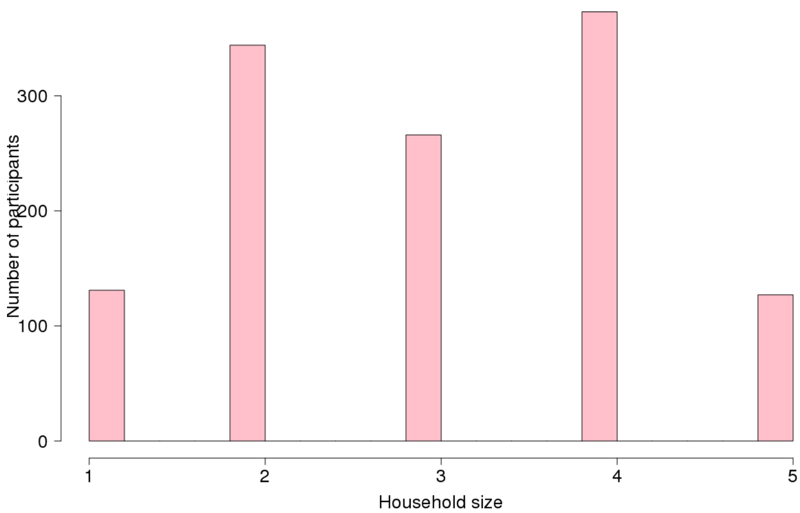
Source: MOBIS-Data

Figure 23: Household size Distribution of the pricing group



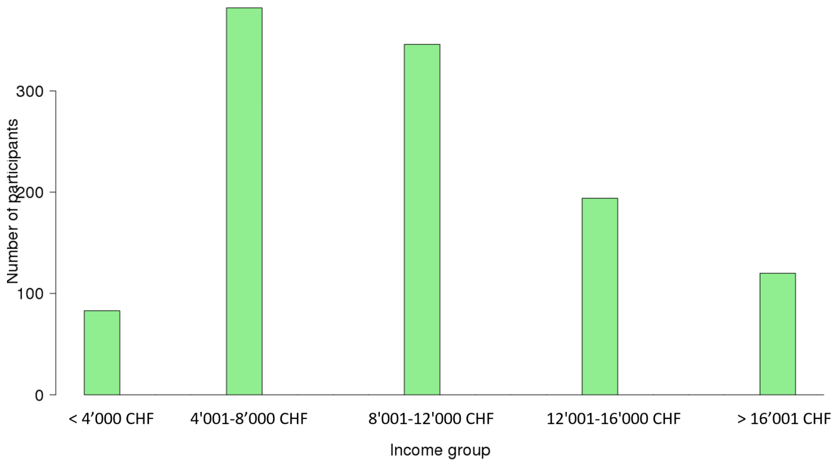
Source: MOBIS-Data

Figure 24: Household size Distribution of the innovation group



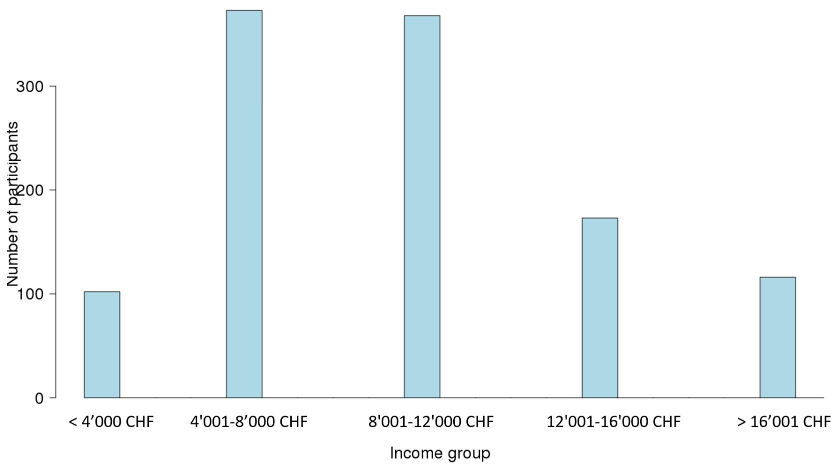
Source: MOBIS-Data

Figure 25: Income Distribution of the control group



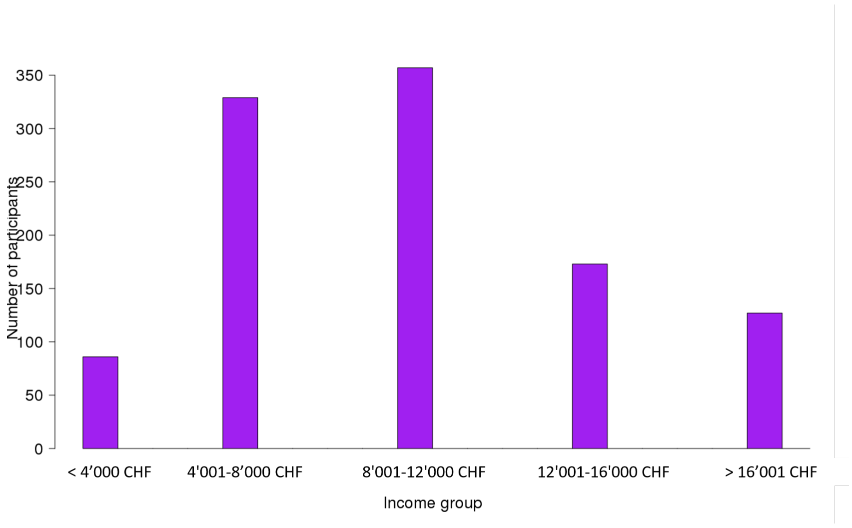
Source: MOBIS-Data

Figure 26: Income Distribution of the nudging group



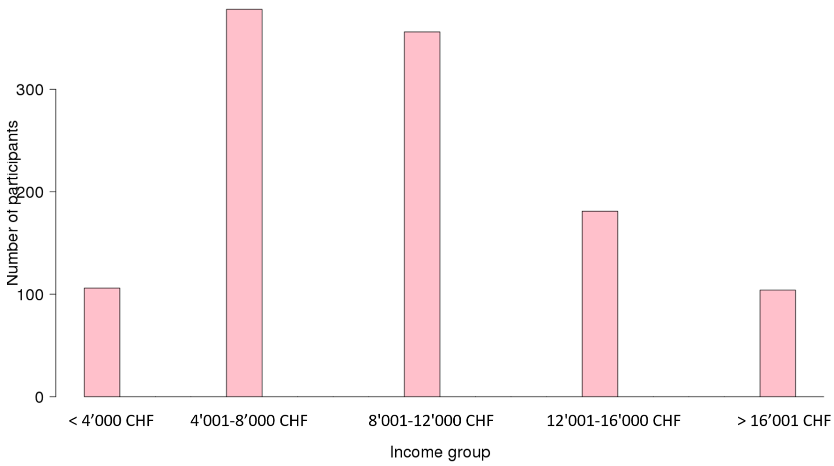
Source: MOBIS-Data

Figure 27: Income Distribution of the pricing group



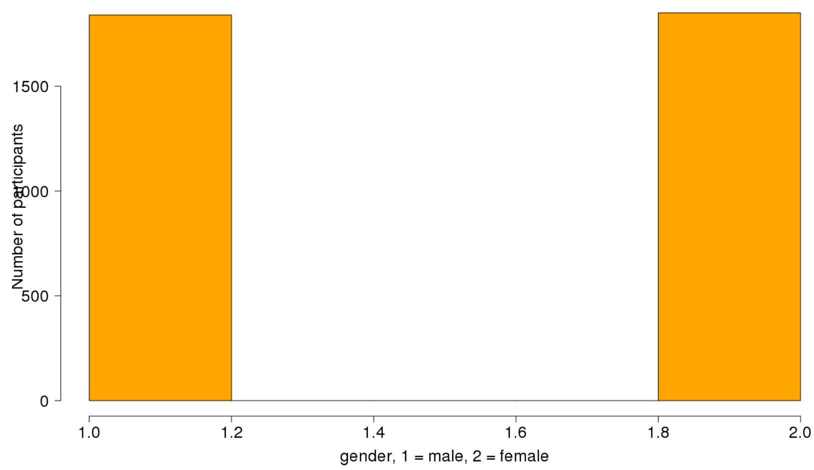
Source: MOBIS-Data

Figure 28: Income Distribution of the innovation group



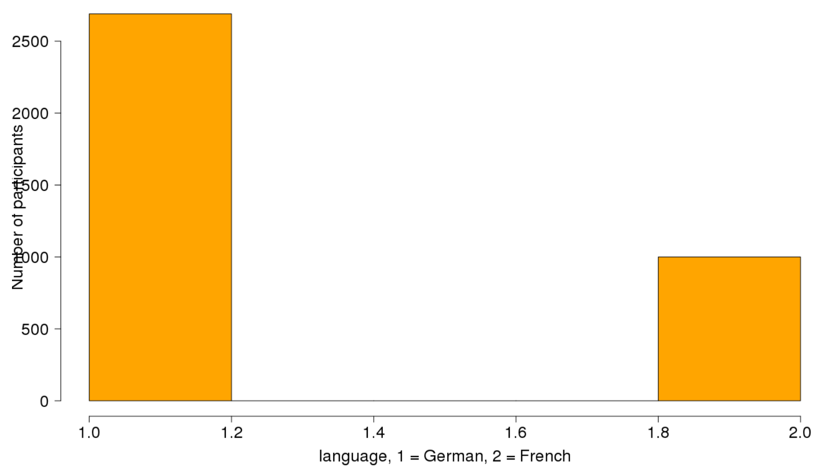
Source: MOBIS-Data

Figure 29: Gender Distribution of all participants



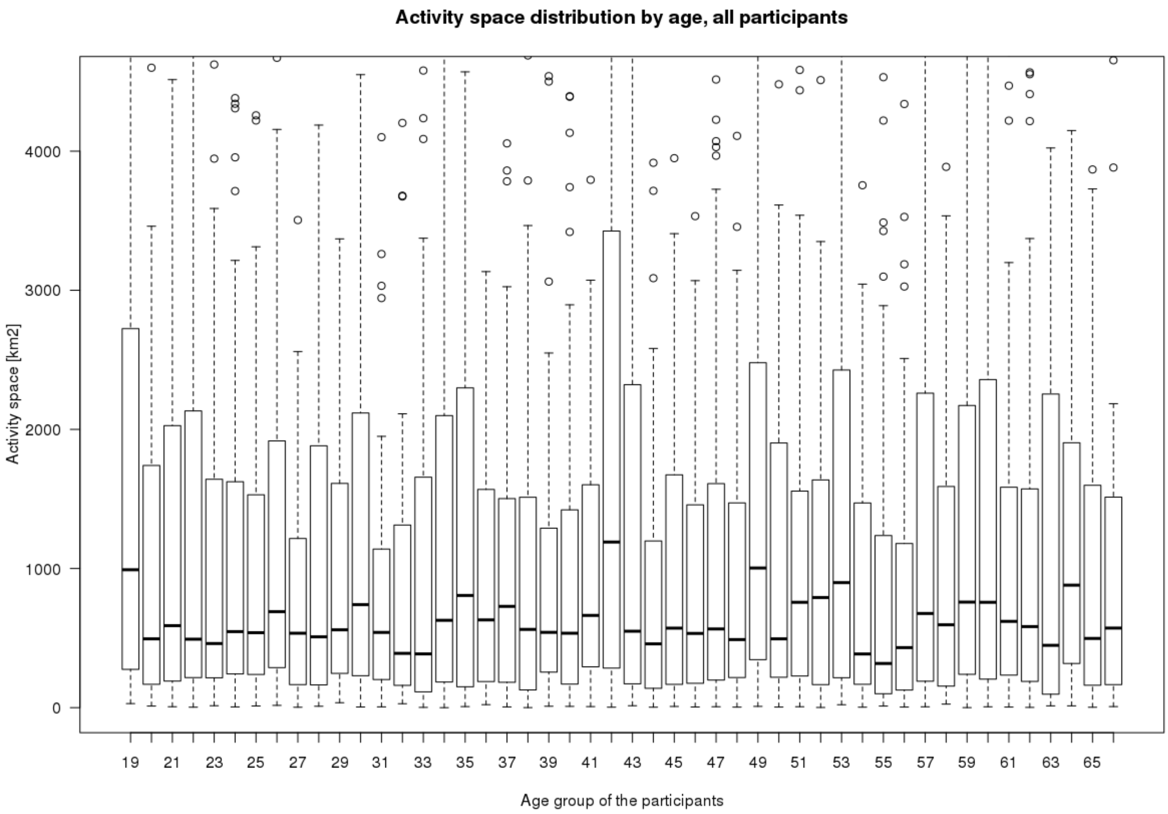
Source: MOBIS-Data

Figure 30: Language Distribution of all participants



Source: MOBIS-Data

Figure 31: Activity space distribution by age



Source: MOBIS-Data

B R-Code

B.1 R-Code: Participants


```
#### Participants
library(readr)
library(dplyr)
library(ggmap)
library(ggplot2)
library(lubridate)


load("/data/mobis/data/mobis_r_workspace.RData")
# mobis_activities_df: activities and accessability of all participants
# mobis_legs_df
# participants: all participants
# activities <- read_csv("/data/mobis/data/csv/activities.csv") old file
codebook_questions <- read_csv("/data/mobis/data/csv/codebook/codebook_questions.csv")
codebook_answers <- read_csv("/data/mobis/data/csv/codebook/codebook_answers.csv")

my_activities_8weeks <- mobis_activities_df %>%
  filter(treatment=='Control') %>%
  filter(phase <= 2) %>%
  filter(in_switzerland == TRUE)


############################################## new grouping
teilnehmer <- participants %>%
  select(participant_ID, postcode_home, household_size, education, age, income, language,
gender, treatment_group, completed_study) %>%
  filter(completed_study == '1')

teilnehmer_nudging <- teilnehmer %>%
  filter(treatment_group == 'Nudging')
teilnehmer_pricing <- teilnehmer %>%
  filter(treatment_group == 'Pricing')
teilnehmer_control <- teilnehmer %>%
  filter(treatment_group == 'Control')
total_participants <- teilnehmer


## Observation period, determine how long each user was tracking his activities
######
teilnehmer_duration <- teilnehmer %>%
  filter(participant_ID %in% unique(mobis_activities_df$user_id))
Activities_observtime <- mobis_activities_df %>%
  group_by(user_id, day = floor_date(started_at, 'day')) %>%
  select(user_id, treatment, started_at, day) %>%
  ungroup()
Activities_observtime_num_days <- Activities_observtime%>%
  group_by(user_id, day) %>%
  #mutate(numb = 1) %>%
  summarise(day_of_obs = mean(day))%>%
  slice(1:5000)
tags <- Activities_observtime_num_days %>%
  group_by(user_id, day) %>%
  slice(1:10000)%>%
  summarise(date = sum(numb)) #%>%
  ungroup()#%>%
  summarize(new_activities = sum(is_new_activity_location_on_day))

teilnehmer_innovation_csv <-
read_csv("~/mobis_analysis/r/eth_activity_spaces/teilnehmer_innovation.csv")
teilnehmer_innovation <- total_participants %>%
  filter(participant_ID %in% unique(teilnehmer_innovation_csv$user_id))



#####
## First version - Cleaning Participants data
##########
my_participants_8week <- participants %>%
  filter(participant_ID %in% unique(my_activities_8weeks$user_id))
```

```
my_participants_8weeks_extract <- my_participants_8week %>%
  select(participant_ID, postcode_home, household_size, education, age, income, language, sex,
         citizen_1, citizen_2, citizen_3, work_type, workload_job,
         employment_1, employment_2, employment_3, work_postcode_1, work_postcode_2,
         work_percentage_1, work_percentage_2,
         capable_walk, smartphone, professional_driver, completed_study)

my_participants_8weeks_male <- my_participants_8weeks_extract %>%
  filter(sex == 1)
my_participants_8weeks_female <- my_participants_8weeks_extract %>%
  filter(sex == 2)

###### getting the 4 week participants
my_participants_4week <- participants %>%
  filter(participant_ID %in% unique(my_activities_4weeks$user_id))

my_participants_4weeks_extract <- my_participants_4week %>%
  select(participant_ID, postcode_home, household_size, education, age, income, language, sex,
         citizen_1, citizen_2, citizen_3)


#####
## Plotting values for my_participants_8week
####################################################
# AGE
#####
hist(teilnehmer_control$age, breaks = 25, col = 'lightgreen',
     cex.main = 1.5, cex.axis = 1.65, cex.lab = 1.65,
     main = '',
     xlab = 'Age of the participants', ylab = 'Number of participants', las = 1)

hist(teilnehmer_nudging$age, breaks = 25, col = 'lightblue',
     cex.main = 1.5, cex.axis = 1.65, cex.lab = 1.65,
     main = '',
     xlab = 'Age of the participants', ylab = 'Number of participants', las = 1)

hist(teilnehmer_pricing$age, breaks = 25, col = 'purple',
     cex.main = 1.5, cex.axis = 1.65, cex.lab = 1.65,
     main = '',
     xlab = 'Age of the participants', ylab = 'Number of participants', las = 1)

hist(total_participants$age, breaks = 25, col = 'orange',
     cex.main = 1.5, cex.axis = 1.65, cex.lab = 1.65,
     main = '',
     xlab = 'Age of the participants', ylab = 'Number of participants', las = 1)

hist(teilnehmer_innovation$age, breaks = 25, col = 'pink',
     cex.main = 1.5, cex.axis = 1.65, cex.lab = 1.65,
     main = '',
     xlab = 'Age of the participants', ylab = 'Number of participants', las = 1)

#####
# HOUSEHOLD SIZE
#####
hist(teilnehmer_control$household_size)
hist(teilnehmer_control$household_size, breaks = 15, col = 'lightgreen',
     cex.main = 1.5, cex.axis = 1.65, cex.lab = 1.65,
     main = '',
     xlab = 'Household size', ylab = 'Number of participants', las = 1)

hist(teilnehmer_nudging$household_size, breaks = 25, col = 'lightblue',
     cex.main = 1.5, cex.axis = 1.65, cex.lab = 1.65,
     main = '',
     xlab = 'Household size', ylab = 'Number of participants', las = 1)

hist(teilnehmer_pricing$household_size, breaks = 25, col = 'purple',
     cex.main = 1.5, cex.axis = 1.65, cex.lab = 1.65,
     main = '',
```

```
     xlab = 'Household size', ylab = 'Number of participants', las = 1)

hist(total_participants$household_size, breaks = 25, col = 'orange',
     cex.main = 1.5, cex.axis = 1.65, cex.lab = 1.65,
     main = '',
     xlab = 'Household size', ylab = 'Number of participants', las = 1)

hist(teilnehmer_innovation$household_size, breaks = 25, col = 'pink',
     cex.main = 1.5, cex.axis = 1.65, cex.lab = 1.65,
     main = '',
     xlab = 'Household size', ylab = 'Number of participants', las = 1)

#####
# Income
#####
income_group <- c('<4000','4001-8000', '8001-12000', '12001-16000', '>16001')
hist(teilnehmer_control$income, breaks = 500, col = 'lightgreen',
     cex.main = 1.5, cex.axis = 1.65, cex.lab = 1.65,
     main = '',
     xlab = 'Income group', ylab = 'Number of participants', las = 1,
     xlim = c(1,5))

hist(teilnehmer_nudging$income, breaks = 500, col = 'lightblue',
     cex.main = 1.5, cex.axis = 1.65, cex.lab = 1.65,
     main = '',
     xlab = 'Income group', ylab = 'Number of participants', las = 1,
     xlim = c(1,5))

hist(teilnehmer_pricing$income, breaks = 500, col = 'purple',
     cex.main = 1.5, cex.axis = 1.65, cex.lab = 1.65,
     main = '',
     xlab = 'Income group', ylab = 'Number of participants', las = 1,
     xlim = c(1,5))

hist(total_participants$income, breaks = 500, col = 'orange',
     cex.main = 1.5, cex.axis = 1.65, cex.lab = 1.65,
     main = '',
     xlab = 'Income group', ylab = 'Number of participants', las = 1,
     xlim = c(1,5))

hist(teilnehmer_innovation$income, breaks = 500, col = 'pink',
     cex.main = 1.5, cex.axis = 1.65, cex.lab = 1.65,
     main = '',
     xlab = 'Income group', ylab = 'Number of participants', las = 1,
     xlim = c(1,5))

#####
# changing values into text form the codebook
participants_codebook_change <- participants %>%
  mutate(education = ifelse(participants$education == 1, "Mandatory education",
                           ifelse(participants$education == 2,"Secondary education (e.g.,
apprenticeship or diploma)",
                                  "Higher education (e.g., university)"))) %>%
  mutate(language = ifelse(participants$language == 'de', 1,
                         ifelse(participants$language == 'fr', 2, 3))) %>%
  mutate(income = )
hist(participants_codebook_change$language)

#####
# gender and language
participants_codebook_change <- total_participants %>%
  mutate(gender = ifelse(total_participants$gender == 'female', 2, 1)) %>%
  mutate(language = ifelse(total_participants$language == 'de', 1,
                         ifelse(total_participants$language == 'fr', 2, 3)))


hist(participants_codebook_change$gender, breaks = 4, col = 'orange',
     cex.main = 1.5, cex.axis = 1.65, cex.lab = 1.65,
     main = '',
```

```
    xlab = 'gender, 1 = male, 2 = female', ylab = 'Number of participants', las = 1,
    xlim = c(1,2))

hist(participants_codebook_change$language, breaks = 4, col = 'orange',
    cex.main = 1.5, cex.axis = 1.65, cex.lab = 1.65,
    main = '',
    xlab = 'language, 1 = German, 2 = French', ylab = 'Number of participants', las = 1)
```

B.2 R-Code: Activity Space for groups


```
### activity spaces for the groups of participants

library(dplyr)
library(sf)
library(car)
library(ggmap)
library(ggplot2)
library(lubridate)
library(sp)

load("/data/mobis/data/mobis_r_workspace.RData")
# mobis_activities_df: activities and accessability of all participants
# mobis_legs_df
# participants: all participants
my_activities_8weeks <- mobis_activities_df %>%
  filter(treatment=='Control') %>%
  filter(phase <= 2) %>%
  filter(in_switzerland == TRUE)

pricing_activities <- mobis_activities_df %>%
  filter(user_id %in% unique(teilnehmer_pricing$participant_ID)) %>%
  filter(phase <= 2) %>%
  filter(in_switzerland == TRUE)
nudging_activities <- mobis_activities_df %>%
  filter(user_id %in% unique(teilnehmer_nudging$participant_ID)) %>%
  filter(phase <= 2) %>%
  filter(in_switzerland == TRUE)
control_activities <- mobis_activities_df %>%
  filter(user_id %in% unique(teilnehmer_control$participant_ID)) %>%
  filter(phase <= 2) %>%
  filter(in_switzerland == TRUE)
innovation_activities <- mobis_activities_df %>%
  filter(user_id %in% unique(teilnehmer_innovation$participant_ID)) %>%
  filter(phase <= 2) %>%
  filter(in_switzerland == TRUE)

## coordinate structure, change into reference system EPSG:2056 to calculate ellipses
my_activities_sdf_pricing <- st_sf(pricing_activities)
my_activities_sdf_pricing$geometry <- st_transform(my_activities_sdf_pricing$geometry, crs =
2056)
my_activities_sdf_nudging <- st_sf(nudging_activities)
my_activities_sdf_nudging$geometry <- st_transform(my_activities_sdf_nudging$geometry, crs =
2056)
my_activities_sdf_control <- st_sf(control_activities)
my_activities_sdf_control$geometry <- st_transform(my_activities_sdf_control$geometry, crs =
2056)
my_activities_sdf_innovation <- st_sf(innovation_activities)
my_activities_sdf_innovation$geometry <- st_transform(my_activities_sdf_innovation$geometry,
crs = 2056)

all_activities <- bind_rows(my_activities_sdf_pricing, my_activities_sdf_nudging,
my_activities_sdf_control)

#####
#Not used
####### group participants by gender
my_activities_sdf_8weeks_male <- my_activities_sdf_8weeks %>%
  filter(user_id %in% unique(my_participants_8weeks_male$participant_ID))
my_activities_sdf_8weeks_female <- my_activities_sdf_8weeks %>%
  filter(user_id %in% unique(my_participants_8weeks_female$participant_ID))

# ploting the activity locations
# plot the point in which an activity is done
plot(my_activities_sdf_8weeks$geometry[my_activities_sdf_8weeks$user_id=='AACZJ'])# if only one
user should be shown, define it here
plot(my_activities_sdf_8weeks$geometry) #all the geometry data

# Computations for only one selected user
#####
```

```
#### Computations for only one selected user. This shows examplarery how the activity space for a distinct user looks
#select a user you want to analyse
mean(my_activity_spaces_df_8weeks$ellipse_area) # MCNDV; determine what should be plotted
median((my_activity_spaces_df_8weeks$ellipse_area)) #ZAVBL
user <- 'MCNDV'
# define the data you want to plot
my_activities_8weeks_user <- my_activities_sdf_control %>%
  filter(user_id == user)


# ploting ellipses for a specific participant
coordinates <- st_coordinates(my_activities_8weeks_user$geometry) # for other user, change the user_id
centroid = colMeans(coordinates)
vcov_matrix = var(coordinates)

# or weighted by duration
weighted_cov = cov.wt(coordinates, my_activities_8weeks_user$duration)
w_centroid <- weighted_cov$center
w_vcov_matrix <- weighted_cov$cov

# calculate the ellipse geometry
ellipse_geometry <- car::ellipse(center = centroid, shape = vcov_matrix,radius = sqrt(qchisq(.95, df=2)), draw=F)
w_ellipse_geometry <- car::ellipse(center = w_centroid, shape = w_vcov_matrix,radius = sqrt(qchisq(.95, df=2)), draw=F)

#convert this to an sf object so that we can plot it
el <- st_sfc(st_polygon(list(ellipse_geometry)), crs=st_crs(my_activities_sdf_8weeks))
w_el <- st_sfc(st_polygon(list(w_ellipse_geometry)), crs=st_crs(my_activities_sdf_8weeks))

# plot the created confidence_ellipse (blue is unweighted, green is weighted)
ggplot(my_activities_8weeks_user, datum=st_crs(4326)) +
  geom_sf() +
  geom_sf(data=el, alpha=0.2, fill='lightblue') +
  geom_sf(data=w_el, alpha=0.2, fill='lightgreen') +
  coord_sf(crs=st_crs(2056), datum = sf::st_crs(2056))

plot(coordinates, xlab = 'longitude', ylab = 'latitude', xlim = (c(2580000, 2660000)), ylim = c(1130000, 1270000))#,
     #main = 'Confidence ellipse for user MCNDV') #plain coordinate system with activities
car::ellipse(center = centroid, shape = vcov_matrix,radius = sqrt(qchisq(.95, df=2)),col = 'darkgreen', size = 150) #plot unweighted ellipse around it
car::ellipse(center = w_centroid, shape = w_vcov_matrix,radius = sqrt(qchisq(.95, df=2)),col = 'darkblue', size = 150) # plot weighted ellipse

#What is the area of the ellipse?
area_95_ci_km <- pi * 5.991* sqrt(prod(eigen(vcov_matrix)$values)) / 1000^2
area_95_ci_km
w_area_95_ci_km <- pi * 5.991* sqrt(prod(eigen(w_vcov_matrix)$values)) / 1000^2
w_area_95_ci_km

#Difference between weighted and unweighted ellipses
ratio_area_ellipse <- w_area_95_ci_km/area_95_ci_km
ratio_area_ellipse

## do get a better representation for the plots, depict them within the scope of switzerland
# load the Kantons shapefile
kantons <-
sf::st_read('/data/mobis/data/geodata/ch_boundaries/swissBOUNDARIES3D_1_3_TLM_KANTONSGEBIET.shp')
 %>% st_set_crs(2056)

# do the ellipse again
coordinates <- st_coordinates(my_activities_8weeks_user$geometry)
weighted_cov = cov.wt(coordinates, my_activities_8weeks_user$duration)
w_centroid <- weighted_cov$center
w_vcov_matrix <- weighted_cov$cov
```

```
# plot activity space for a particular user
ggplot() + #you can organise you code nicely by putting the ggplot layers on new lines -  the
'+' symbol links the code together
  geom_sf(data = kantons, color = 'black') +
  geom_sf(data = my_activities_8weeks_user,
          inherit.aes = FALSE,
          colour = "#238443",
          fill = "#238443",
          alpha = 0.5,
          size = 6,
          shape = 21) +
  geom_sf(data = w_el, alpha = 0.2, fill = 'blue') +
  coord_sf(crs = st_crs(2056), datum = sf::st_crs(2056)) + #important to display the map
properly - try it without it.
  labs(title = 'Activity space and locations for user MCNDV')# , caption = 'Notice how some are
outside of switzerland')

# activity space for a particular user grouped in a specific time frame, e.g days or weeks
calculate_ellipse_area_user <- function(df1) {
  coordinates <- st_coordinates(df1$geometry)
  weighted_cov = cov.wt(coordinates, df1$duration)
  w_centroid <- weighted_cov$center
  w_vcov_matrix <- weighted_cov$cov

  p <-  ifelse(
    anyNA(w_vcov_matrix), NA, prod(eigen(w_vcov_matrix)$values)
  )
  ellipse_radius = ifelse(is.na(p) || p < 0, NA, sqrt(p))
  area <- pi * 5.991* ellipse_radius / 1000^2

  el <- car::ellipse(center = w_centroid, shape = w_vcov_matrix,radius = sqrt(qchisq(.95,
df=2)), draw = F)
  el_geom <- st_sfc(st_polygon(list(el)), crs=st_crs(my_activities_8weeks_user))
  return (data.frame(
    centroid_X = w_centroid['X'], centroid_Y = w_centroid['Y'],
    cov_XX = w_vcov_matrix[1,1], cov_XY = w_vcov_matrix[1,2], cov_YX = w_vcov_matrix[2,1],
cov_YY = w_vcov_matrix[2,2],
    ellipse_area=area, stringsAsFactors=F,
    geometry=el_geom))
}

# plot the activity spaces for a partucular user in daily resolution and draw the linear
regression line
area_ellipse_user <- my_activities_8weeks_user %>%
  mutate(date = floor_date(started_at,'week')) %>%
  group_by(user_id, date) %>%
  mutate(n=n()) %>%
  filter(n>=3) %>%
  group_modify(~calculate_ellipse_area_user(.x), keep = TRUE) %>%
  st_as_sf(crs=st_crs(my_activities_8weeks_user)) %>% #need this line to put the projection
(2056) back
  ungroup()
area_ellipse_user
summary(area_ellipse_user)
plot(area_ellipse_user$geometry)
plot(area_ellipse_user$date, area_ellipse_user$ellipse_area)
regModel <- lm(ellipse_area ~ date, data = area_ellipse_user)
regModel
abline(regModel$coef[1], regModel$coef[2], col = 'red', lwd = 3)
summary(regModel)


#####
# calculate the numbers for the whole data set which should be analysed. in this section the
'control' group participants with a time frame of 8 weeks of observation are calculated here
#####
calculate_ellipse_area <- function(df1) {
  coordinates <- st_coordinates(df1$geometry)
  weighted_cov = cov.wt(coordinates, df1$duration)
```

```
  w_centroid <- weighted_cov$center
  w_vcov_matrix <- weighted_cov$cov

  p <-  ifelse(
    anyNA(w_vcov_matrix), NA, prod(eigen(w_vcov_matrix)$values)
  )
  ellipse_radius = ifelse(is.na(p) || p < 0, NA, sqrt(p))
  area <- pi * 5.991* ellipse_radius / 1000^2

  el <- car::ellipse(center = w_centroid, shape = w_vcov_matrix,radius = sqrt(qchisq(.95,
df=2)), draw = F)
  mode(el) <- "integer"
  aa <- list(el)
  el_geom <- st_sfc(st_polygon(list(el)), crs=st_crs(my_activities_sdf_group))
  return (data.frame(
    centroid_X = w_centroid['X'], centroid_Y = w_centroid['Y'],
    cov_XX = w_vcov_matrix[1,1], cov_XY = w_vcov_matrix[1,2], cov_YX = w_vcov_matrix[2,1],
cov_YY = w_vcov_matrix[2,2],
    ellipse_area=area, stringsAsFactors=F,
    geometry=el_geom))
}

#####
# Which group? Pricing, control, nudging,innovation all?
#####
my_activities_sdf_group <- my_activities_sdf_pricing
participants_group <- teilnehmer_pricing

# calculate daily num activities and ellipse area for each person, covid vs mobis
my_activity_spaces_df_group <- my_activities_sdf_group %>%
  #filter(user_id  %in% unique(my_activities_8weeks$user_id)) %>%   # [1:2]
  #slice(1:3000) %>% # do this first to only work with a small sample to test
  #filter(user_id %in% sample_activities)  %>% ### maybe do some fieltering here
  # mutate(date = floor_date(started_at,'week')) %>% ## for splitting sample into time period
  group_by(user_id) %>%
  mutate(n=n()) %>%
  filter(n>=3) %>%
  group_modify(~ calculate_ellipse_area(.x), keep = TRUE) %>% #this is how you use the function
  st_as_sf(crs=st_crs(my_activities_8weeks)) %>% #need this line to put the projection (2056)
back
  ungroup()

#summary(my_activity_spaces_df_group$ellipse_area)
#plot(my_activity_spaces_df_group$date, my_activity_spaces_df_group$ellipse_area)
#regModel <- lm(ellipse_area ~ date, data = my_activity_spaces_df_group)
#regModel
#abline(regModel$coef[1], regModel$coef[2], col = 'red', lwd = 3)
#summary(regModel)

#####
#display all the home locations, and two activty spaces
#####
ggplot() + #you can organise you code nicely by putting the ggplot layers on new lines -  the
'+' symbol links the code together
  geom_sf(data = kantons, color = 'black')  +
  geom_sf(data = my_activities_sdf_group %>% filter(in_switzerland), alpha=0.1) +
  geom_sf(data = my_activity_spaces_df_group,
          inherit.aes = FALSE,
          colour = "#238443",
          fill = "#238443",
          alpha = 0.1,
          shape = 21)+
  coord_sf(crs=st_crs(2056), datum = sf::st_crs(2056)) + #important to display the map properly
- try it without it.
  labs(title='Activity spaces for users, with all activities mapped at 10% opacity',
caption='')

#####
#getting plot data from participants and activity spaces
```

```
# getting the same length of argument
participants_for_analysise_group <- participants_group %>%
  filter(participant_ID %in% unique(my_activity_spaces_df_group$user_id))
my_activity_spaces_df_for_analysise_group <- my_activity_spaces_df_group %>%
  filter(user_id %in% unique(participants_group$participant_ID))

participants_activity_df_group <- dplyr::bind_cols(participants_for_analysise_group,
my_activity_spaces_df_for_analysise_group %>% select(-user_id))
participants_activity_df_group


#####
# key figures groups
#####
# control
#####
length_control <- nrow(participants_activity_df_group)
mean_control <- mean(participants_activity_df_group$ellipse_area)
median_control <- median(participants_activity_df_group$ellipse_area)
sd_control <-  sd(participants_activity_df_group$ellipse_area)
quantile1_control <- quantile(participants_activity_df_group$ellipse_area, 0.25)
quantile2_control <- quantile(participants_activity_df_group$ellipse_area, 0.75)

key_figures_control <- c(length_control, mean_control, median_control, sd_control,
quantile1_control, quantile2_control)
key_figures_control

age_control <- key_figures_age_group_control
household_control <- key_figures_household_group_control
income_control <- key_figures_income_group_control
language_control <- key_figures_language_group_control

label <- c('length', 'mean', 'std.mean', 'median', 'std. median')
key_figures_df_control <- data.frame(label, age_control, household_control,
                                    income_control,language_control)
key_figures_df_control



#####
# nudging
#####
length_nudging <- nrow(participants_activity_df_group)
mean_nudging <- mean(participants_activity_df_group$ellipse_area)
median_nudging <- median(participants_activity_df_group$ellipse_area)
sd_nudging <-  sd(participants_activity_df_group$ellipse_area)
quantile1_nudging <- quantile(participants_activity_df_group$ellipse_area, 0.25)
quantile2_nudging <- quantile(participants_activity_df_group$ellipse_area, 0.75)

key_figures_nudging <- c(length_nudging, mean_nudging, median_nudging, sd_nudging,
quantile1_nudging, quantile2_nudging)
key_figures_nudging

age_nudging <- key_figures_age_group_nudging
household_nudging <- key_figures_household_group_nudging
income_nudging <- key_figures_income_group_nudging
language_nudging <- key_figures_language_group_nudging

label <- c('length', 'mean', 'std.mean', 'median', 'std. median')
key_figures_df_nudging <- data.frame(label, age_nudging, household_nudging,
                                     income_nudging,language_nudging)
key_figures_df_nudging



#####
# pricing
#####
length_pricing <- nrow(participants_activity_df_group)
```

```
mean_pricing <- mean(participants_activity_df_group$ellipse_area)
median_pricing <- median(participants_activity_df_group$ellipse_area)
sd_pricing <-  sd(participants_activity_df_group$ellipse_area)
quantile1_pricing <- quantile(participants_activity_df_group$ellipse_area, 0.25)
quantile2_pricing <- quantile(participants_activity_df_group$ellipse_area, 0.75)

key_figures_pricing <- c(length_pricing, mean_pricing, median_pricing, sd_pricing,
quantile1_pricing, quantile2_pricing)
key_figures_pricing

age_pricing <- key_figures_age_group_pricing
household_pricing <- key_figures_household_group_pricing
income_pricing <- key_figures_income_group_pricing
language_pricing <- key_figures_language_group_pricing

label <- c('length', 'mean', 'std.mean', 'median', 'std. median')
key_figures_df_pricing <- data.frame(label, age_pricing, household_pricing,
                                     income_pricing,language_pricing)
key_figures_df_pricing


#####
# Innovation
#####
length_innovation <- nrow(participants_activity_df_group)
mean_innovation <- mean(participants_activity_df_group$ellipse_area)
median_innovation <- median(participants_activity_df_group$ellipse_area)
sd_innovation <-  sd(participants_activity_df_group$ellipse_area)
quantile1_innovation <- quantile(participants_activity_df_group$ellipse_area, 0.25)
quantile2_innovation <- quantile(participants_activity_df_group$ellipse_area, 0.75)

key_figures_innovation <- c(length_innovation, mean_innovation, median_innovation,
sd_innovation, quantile1_innovation, quantile2_innovation)
key_figures_innovation

age_innovation <- key_figures_age_group_innovation
household_innovation <- key_figures_household_group_innovation
income_innovation <- key_figures_income_group_innovation
language_innovation <- key_figures_language_group_innovation

label <- c('length', 'mean', 'std.mean', 'median', 'std. median')
key_figures_df_innovation <- data.frame(label, age_innovation, household_innovation,
                                     income_innovation,language_innovation)
key_figures_df_innovation

#####
# activity space vs age
#####
participants_activity_df_age_group <- participants_activity_df_group %>%
  filter(!is.na(age))
length_age_group <- nrow(participants_activity_df_age_group)
length_age_group

mean_activity_space_age_group <- participants_activity_df_age_group %>%
  group_by(age) %>%
  summarise(mean_space_age = mean(ellipse_area))
median_activity_space_age_group <- participants_activity_df_age_group %>%
  group_by(age) %>%
  summarise(median_space_age = median(ellipse_area))
sd_activity_space_age_group <- participants_activity_df_age_group %>%
  group_by(age) %>%
  summarise(sd_space_age = sd(ellipse_area))
activity_space_age_group <- bind_cols(mean_activity_space_age_group,
median_activity_space_age_group %>% select(-age), sd_activity_space_age_group%>% select(-age))
activity_space_age_group

participants_activity_df_age_group$AgeGroup <- cut(participants_activity_df_age_group$age,
breaks = c(seq(10, 65, by = 10), Inf))
boxplot(ellipse_area ~ AgeGroup, participants_activity_df_age_group,
```

```
        cex.main = 1.2, cex.axis = 1.0, cex.lab = 1.0,
        main = 'Activity space distribution by age, innovation',
        ylab = 'Activity space [km2]', xlab='Age group of the participants',
        ylim = c(0, 4500), las = 1)

boxplot(ellipse_area ~ age, participants_activity_df_age_group,
        cex.main = 1.4, cex.axis = 1.2, cex.lab = 1.2,
        main = 'Activity space distribution by age, innovation',
        ylab = 'Activity space [km2]', xlab='Age group of the participants',
        ylim = c(0, 4500), las = 1)

mean_age_group <- mean(activity_space_age_group$mean_space_age)
median_age_group <- median(activity_space_age_group$median_space_age)
sd_mean_age_group <- sd(activity_space_age_group$mean_space_age)
sd_median_age_group <-  sd(activity_space_age_group$median_space_age)

key_figures_age_group_control <- c(length_age_group, mean_age_group, sd_mean_age_group,
median_age_group, sd_median_age_group)
key_figures_age_group_control

key_figures_age_group_nudging <- c(length_age_group, mean_age_group, sd_mean_age_group,
median_age_group, sd_median_age_group)
key_figures_age_group_nudging

key_figures_age_group_pricing <- c(length_age_group, mean_age_group, sd_mean_age_group,
median_age_group, sd_median_age_group)
key_figures_age_group_pricing

key_figures_age_group_innovation <- c(length_age_group, mean_age_group, sd_mean_age_group,
median_age_group, sd_median_age_group)
key_figures_age_group_innovation

#ggplot(activity_space_age, aes(age, y = value, color = variable)) +
#  geom_point(aes(y = mean_space_age, col = "blue")) +
#  geom_line(aes(y = median_space_age, col = "red"))+
#  labs(title = "Distribution of the activity spaces by age, control group", x = "age of the
participants", y = "Activity Space area [km^2]", color = "") +
#  scale_color_manual(labels = c("mean", "median"), values = c("blue", "red"))




#####
# household_size
#####
participants_activity_df_household_group <- participants_activity_df_group %>%
  filter(!is.na(household_size))
length_household_group <- nrow(participants_activity_df_household_group)
length_household_group

mean_activity_space_household_group <- participants_activity_df_household_group %>%
  group_by(household_size) %>%
  summarise(mean_space_household = mean(ellipse_area))
median_activity_space_household_group <- participants_activity_df_household_group %>%
  group_by(household_size) %>%
  summarise(median_space_household = median(ellipse_area))
sd_activity_space_household_group <- participants_activity_df_household_group %>%
  group_by(household_size) %>%
  summarise(sd_space_household = sd(ellipse_area))
activity_space_household_group <- bind_cols(mean_activity_space_household_group,
median_activity_space_household_group %>% select(-household_size),
                                            sd_activity_space_household_group %>% select(-
household_size))
activity_space_household_group

boxplot(ellipse_area ~ household_size, participants_activity_df_household_group,
        cex.main = 1.2, cex.axis = 1.1, cex.lab = 1.2,
        main = 'Activity space distribution by household size, innovation',
        ylab = 'Activity space [km2]', xlab='household size',
```

```
        ylim = c(0, 4500), las = 1)

mean_household_group <- mean(activity_space_household_group$mean_space_household)
median_household_group <- median(activity_space_household_group$median_space_household)
sd_mean_household_group <- sd(activity_space_household_group$mean_space_household)
sd_median_household_group <-  sd(activity_space_household_group$median_space_household)

key_figures_household_group_control <- c(length_household_group, mean_household_group,
sd_mean_household_group, median_household_group, sd_median_household_group)
key_figures_household_group_control

key_figures_household_group_nudging <- c(length_household_group, mean_household_group,
sd_mean_household_group, median_household_group, sd_median_household_group)
key_figures_household_group_nudging

key_figures_household_group_pricing <- c(length_household_group, mean_household_group,
sd_mean_household_group, median_household_group, sd_median_household_group)
key_figures_household_group_pricing

key_figures_household_group_innovation <- c(length_household_group, mean_household_group,
sd_mean_household_group, median_household_group, sd_median_household_group)
key_figures_household_group_innovation

#ggplot(activity_space_household, aes(household_size, y = value, color = variable)) +
#  geom_point(aes(y = mean_space_household, col = "blue"), size = 3) +
#  geom_point(aes(y = median_space_household, col = "red"), size = 3)+
#  labs(title = "Distribution of activity space area by household size, control group", x =
"household size", y = "Activity space area [km2]", color = "variable") +
#  scale_color_manual(labels = c("mean", "median"), values = c("blue", "red"))



#####
# activity space vs income
#####
participants_activity_df_income_group <- participants_activity_df_group %>%
  filter(income != 99)
length_income_group <- nrow(participants_activity_df_income_group)
length_income_group

mean_activity_space_income_group <- participants_activity_df_income_group %>%
  group_by(income) %>%
  summarise(mean_space_income = mean(ellipse_area))
median_activity_space_income_group <- participants_activity_df_income_group %>%
  group_by(income) %>%
  summarise(median_space_income = median(ellipse_area))
sd_activity_space_income_group <- participants_activity_df_income_group %>%
  group_by(income) %>%
  summarise(sd_space_income = sd(ellipse_area))
activity_space_income_group <- bind_cols(mean_activity_space_income_group,
median_activity_space_income_group %>% select(-income),
                                         sd_activity_space_income_group%>% select(-income))
activity_space_income_group

boxplot(ellipse_area ~ income, participants_activity_df_income_group,
        cex.main = 1.2, cex.axis = 1.1, cex.lab = 1.2,
        main = 'Activity space distribution by income group, innovation',
        ylab = 'Activity space [km2]', xlab='income group ',
        ylim = c(0, 4000), las = 1)

mean_income_group <- mean(activity_space_income_group$mean_space_income)
median_income_group <- median(activity_space_income_group$median_space_income)
sd_mean_income_group <- sd(activity_space_income_group$mean_space_income)
sd_median_income_group <-  sd(activity_space_income_group$median_space_income)

key_figures_income_group_control <- c(length_income_group, mean_income_group,
sd_mean_income_group, median_income_group, sd_median_income_group)
key_figures_income_group_control
```

```
key_figures_income_group_nudging <- c(length_income_group, mean_income_group,
sd_mean_income_group, median_income_group, sd_median_income_group)
key_figures_income_group_nudging

key_figures_income_group_pricing <- c(length_income_group, mean_income_group,
sd_mean_income_group, median_income_group, sd_median_income_group)
key_figures_income_group_pricing

key_figures_income_group_innovation <- c(length_income_group, mean_income_group,
sd_mean_income_group, median_income_group, sd_median_income_group)
key_figures_income_group_innovation


#plot(activity_space_income$income, activity_space_income$mean_space_income)
#income_label <- c('<4´000', '4´001-8´000', '8´001-12´000', '12´001-16´000', '>16´000')
#ggplot(activity_space_income, aes(income, y = value, color = variable)) +
#  geom_point(aes(y = mean_space_income, col = "blue"), size = 3) +
#  geom_point(aes(y = median_space_income, col = "red"), size = 3)+
#  labs(title = "Distribution of activity space area by income, control group", x = "income", y
= "Actvitiy space area [km2]", color = "") +
#  scale_color_manual(labels = c("mean", "median"), values = c("blue", "red"))
#summary(participants_activity_df_filtered)


######
# gender
######
participants_activity_df_sex_group <- participants_activity_df_group %>%
  filter(!is.na(sex)) %>%
  mutate(sex = ifelse(sex == 1, "Male", "Female"))
length_sex_group <- nrow(participants_activity_df_sex_group)
length_sex_group

boxplot(ellipse_area ~ sex, participants_activity_df_sex_group,
        cex.main = 1.2, cex.axis = 1.2, cex.lab = 1.4,
        main = 'Activity space distribution by gender, innovation',
        ylab = 'Activity space [km2]', xlab='gender',
        ylim = c(0, 4500), las = 1)

#####
# activity space vs culture/language
#####
participants_activity_df_language_group <- participants_activity_df_group %>%
  filter(!is.na(language))
length_language_group <- nrow(participants_activity_df_language_group)
length_language_group

mean_activity_space_language_group <- participants_activity_df_language_group %>%
  group_by(language) %>%
  summarise(mean_space_language = mean(ellipse_area))
median_activity_space_language_group <- participants_activity_df_language_group %>%
  group_by(language) %>%
  summarise(median_space_language = median(ellipse_area))
sd_activity_space_language_group <- participants_activity_df_language_group %>%
  group_by(language) %>%
  summarise(sd_space_language = sd(ellipse_area))
activity_space_language_group <- bind_cols(mean_activity_space_language_group,
median_activity_space_language_group %>% select(-language),
                                           sd_activity_space_language_group%>% select(-
language))
activity_space_language_group

boxplot(ellipse_area ~ language, participants_activity_df_language,
        cex.main = 1.2, cex.axis = 1.2, cex.lab = 1.4,
        main = 'Activity space distribution by language, nudging',
        ylab = 'Activity space [km2]', xlab='language',
        ylim = c(0, 3500), las = 1)

mean_language_group <- mean(activity_space_language_group$mean_space_language)
```

```
median_language_group <- median(activity_space_language_group$median_space_language)
sd_mean_language_group <- sd(activity_space_language_group$mean_space_language)
sd_median_language_group <-  sd(activity_space_language_group$median_space_language)

key_figures_language_group_control <- c(length_language_group, mean_language_group,
sd_mean_language_group, median_language_group, sd_median_language_group)
key_figures_language_group_control

key_figures_language_group_nudging <- c(length_language_group, mean_language_group,
sd_mean_language_group, median_language_group, sd_median_language_group)
key_figures_language_group_nudging

key_figures_language_group_pricing <- c(length_language_group, mean_language_group,
sd_mean_language_group, median_language_group, sd_median_language_group)
key_figures_language_group_pricing

key_figures_language_group_innovation <- c(length_language_group, mean_language_group,
sd_mean_language_group, median_language_group, sd_median_language_group)
key_figures_language_group_innovation

#plot(activity_space_language$language, activity_space_language$mean_space_language)
#ggplot(activity_space_language, aes(language, y = value, color = variable)) +
#  geom_point(aes(y = mean_space_language, col = "blue"), size = 3) +
#  geom_point(aes(y = median_space_language, col = "red"), size = 3)+
#  labs(title = "Distribution of activity space area by language, control group", x =
"language", y = "Actvitiy space area [km2]", color = "") +
#  scale_color_manual(labels = c("mean", "median"), values = c("blue", "red"))
#summary(participants_activity_df_language)
```

B.3 R-Code: Activity Space for all participants


```
##### activity spaces for all the participants

library(dplyr)
library(sf)
library(car)
library(ggmap)
library(ggplot2)
library(lubridate)
library(sp)

load("/data/mobis/data/mobis_r_workspace.RData")
# mobis_activities_df: activities and accessability of all participants
# mobis_legs_df
# participants: all participants
# take existing data

## coordinate structure, change into reference system EPSG:2056 to calculate ellipses
my_activities_sdf_all <- st_sf(all_activities)
my_activities_sdf_all$geometry <- st_transform(all_activities$geometry, crs = 2056)
my_activities_sdf_all

calculate_ellipse_area <- function(df1) {
  coordinates <- st_coordinates(df1$geometry)
  weighted_cov = cov.wt(coordinates, df1$duration)
  w_centroid <- weighted_cov$center
  w_vcov_matrix <- weighted_cov$cov

  p <-  ifelse(
    anyNA(w_vcov_matrix), NA, prod(eigen(w_vcov_matrix)$values)
  )
  ellipse_radius = ifelse(is.na(p) || p < 0, NA, sqrt(p))
  area <- pi * 5.991* ellipse_radius / 1000^2

  el <- car::ellipse(center = w_centroid, shape = w_vcov_matrix,radius = sqrt(qchisq(.95,
df=2)), draw = F)
  mode(el) <- "integer"
  aa <- list(el)
  el_geom <- st_sfc(st_polygon(list(el)), crs=st_crs(my_activities_sdf_all))
  return (data.frame(
    centroid_X = w_centroid['X'], centroid_Y = w_centroid['Y'],
    cov_XX = w_vcov_matrix[1,1], cov_XY = w_vcov_matrix[1,2], cov_YX = w_vcov_matrix[2,1],
cov_YY = w_vcov_matrix[2,2],
    ellipse_area=area, stringsAsFactors=F,
    geometry=el_geom))
}

# calculate daily num activities and ellipse area for each person, covid vs mobis
my_activity_spaces_df_all <- my_activities_sdf_all %>%
  #filter(user_id  %in% unique(my_activities_8weeks$user_id)) %>%   # [1:2]
  #slice(1:10000) %>% # do this first to only work with a small sample to test
  #filter(user_id %in% sample_activities)  %>% ### maybe do some fieltering here
  # mutate(date = floor_date(started_at,'week')) %>% ## for splitting sample into time period
  group_by(user_id) %>%
  mutate(n=n()) %>%
  filter(n>=3) %>%
  group_modify(~ calculate_ellipse_area(.x), keep = TRUE) %>% #this is how you use the function
  st_as_sf(crs=st_crs(my_activities_sdf_all)) %>% #need this line to put the projection (2056)
back
  ungroup()

#summary(my_activity_spaces_df_all$ellipse_area)
#var_4week <- var(my_activity_spaces_df_all$ellipse_area)# – die Varianz von x
#sd_4week <- sd(my_activity_spaces_df_all$ellipse_area) #– Die Standardabweichung von x
#quantile_4week <- quantile(my_activity_spaces_df_all$ellipse_area,0.25) # – berechnet das 25%
Quantile von x


# getting the same length of argument
participants_for_analysise_all <- total_participants %>%
```

```
  filter(participant_ID %in% unique(my_activity_spaces_df_all$user_id))
my_activity_spaces_df_for_analysise_all <- my_activity_spaces_df_all %>%
  filter(user_id %in% unique(total_participants$participant_ID))

#####
# Save results for all participants, before plotting
#####
participants_activity_df_all <- dplyr::bind_cols(participants_for_analysise_all,
my_activity_spaces_df_for_analysise_all %>% select(-user_id))
participants_activity_df_all$AgeGroup <- cut(participants_activity_df_all$age, breaks =
c(seq(10, 65, by = 10), Inf))
participants_activity_df_all
setwd("~/mobis_analysis/r/eth_activity_spaces") ##Set to where you want your file
write_csv(participants_activity_df_all, 'activity_spaces_results.csv')

participants_activity_df_all$nonGrouping <- cut(participants_activity_df_all$age, breaks =
c(seq(10, 65, by = 100), Inf))
plot2 <- boxplot(ellipse_area ~ nonGrouping, participants_activity_df_all,
        cex.main = 1.4, cex.axis = 1.2, cex.lab = 1.2,
        main = 'Activity space distribution',
        ylab = 'Activity space [km2]', xlab='all participants',
        ylim = c(0, 4500), las = 1, horizontal = TRUE)
hist(participants_activity_df_age_all$ellipse_area , breaks=1000, main="" , xlab="value of the
variable")

hist(participants_activity_df_age_all$age, participants_activity_df_age_all$ellipse_area, xlim
= c(16, 67))
ggplot(participants_activity_df_age_all, aes()) +
  geom_histogram(x = participants_activity_df_age_all$age, y =
participants_activity_df_age_all$ellipse_area)+
  labs(title = '', x = "age group of the participants", y = "Activity Space area [km^2]", color
= "") #+
  scale_color_manual(labels = c("mean", "median"), values = c("blue", "red"))

## histo for ellipse area
ggplot(data=participants_activity_df_age_all, aes(x =
participants_activity_df_age_all$ellipse_area)) +
  geom_histogram(breaks=seq(0, 2800, by=10),
                 alpha = .2, color = 'orange', fill = 'orange') +
  geom_boxplot()+
  coord_cartesian(xlim = c(0, 2800))+
  xlab('ellipse area [km2]') +# for the x axis label
  ylab('Fequency')+
  theme(axis.title.x = element_text(size = 20), axis.text.x = element_text(size = 15)
,axis.title.y = element_text(size=20), axis.text.y = element_text(size = 15) )
  #labs(title="", x="ellipse area [km2]", y="Fequency", size = 50)

#####
# key figures all participants
#####
length_all <- nrow(participants_activity_df_all)
mean_all <- mean(participants_activity_df_all$ellipse_area)
median_all <- median(participants_activity_df_all$ellipse_area)
sd_all <-  sd(participants_activity_df_all$ellipse_area)
quantile1_all <- quantile(participants_activity_df_all$ellipse_area, 0.25)
quantile2_all <- quantile(participants_activity_df_all$ellipse_area, 0.75)

key_figures_all <- c(length_all, mean_all, median_all, sd_all, quantile1_all, quantile2_all)
key_figures_all


#####
# activity space vs age
#####
participants_activity_df_age_all <- participants_activity_df_all %>%
  filter(!is.na(age))
length_age_all <- nrow(participants_activity_df_age_all)
length_age_all
```

```
mean_activity_space_age_all <- participants_activity_df_age_all %>%
  group_by(age) %>%
  #mutate(space_age = c(ellipse_area)) #%>%
  summarise(mean_space_age = mean(ellipse_area))
median_activity_space_age_all <- participants_activity_df_age_all %>%
  group_by(age) %>%
  #mutate(num_new_activities = sum(new_activities))%>%
  summarise(median_space_age = median(ellipse_area))
std_activity_space_age_all <- participants_activity_df_age_all %>%
  group_by(age) %>%
  #mutate(num_new_activities = sum(new_activities))%>%
  summarise(std_space_age = sd(ellipse_area))
activity_space_age_all <- bind_cols(mean_activity_space_age_all, median_activity_space_age_all
%>% select(-age), std_activity_space_age_all %>% select(-age))
activity_space_age_all

#####
# AgeGroup
participants_activity_df_age_all$AgeGroup <- cut(participants_activity_df_age_all$age, breaks =
c(seq(10, 65, by = 10), Inf))
mean_activity_space_age_all <- participants_activity_df_age_all %>%
  group_by(AgeGroup) %>%
  #mutate(space_age = c(ellipse_area)) #%>%
  summarise(mean_space_age = mean(ellipse_area))
median_activity_space_age_all <- participants_activity_df_age_all %>%
  group_by(AgeGroup) %>%
  #mutate(num_new_activities = sum(new_activities))%>%
  summarise(median_space_age = median(ellipse_area))
std_activity_space_age_all <- participants_activity_df_age_all %>%
  group_by(AgeGroup) %>%
  #mutate(num_new_activities = sum(new_activities))%>%
  summarise(std_space_age = sd(ellipse_area))
length_AgeGroup <- participants_activity_df_age_all %>%
  group_by(AgeGroup) %>%
  mutate(rownumberAge = row_number())%>%
  group_by(AgeGroup) %>%
  summarise(length_space_age = max(rownumberAge))
activity_space_age_all <- bind_cols(mean_activity_space_age_all, median_activity_space_age_all
%>% select(-AgeGroup),
                                    std_activity_space_age_all %>% select(-AgeGroup),
length_AgeGroup %>% select(-AgeGroup))
activity_space_age_all
######

participants_activity_df_age_all$AgeGroup <- cut(participants_activity_df_age_all$age, breaks =
c(seq(10, 65, by = 10), Inf))
boxplot(ellipse_area ~ AgeGroup, participants_activity_df_age_all,
        cex.main = 1.2, cex.axis = 1.2, cex.lab = 1.4,
        main = '',
        ylab = 'Activity space [km2]', xlab='Age group of the participants',
        ylim = c(0, 4500), las = 1,
        par(mar = c(4, 6, 1, 1)))

op <- par(mar = c(5,4,4,2) + 0.1) ## default is c(5,4,4,2) + 0.1
par(op)

boxplot(ellipse_area ~ age, participants_activity_df_age_all,
        cex.main = 1.2, cex.axis = 1.0, cex.lab = 1.0,
        main = 'Activity space distribution by age, all participants',
        ylab = 'Activity space [km2]', xlab='Age group of the participants',
        ylim = c(0, 4500), las = 1)

mean_age_all <- mean(activity_space_age_all$mean_space_age)
median_age_all <- median(activity_space_age_all$median_space_age)
sd_mean_age_all <- sd(activity_space_age_all$mean_space_age)
sd_median_age_all <-  sd(activity_space_age_all$median_space_age)

key_figures_age_all <- c(length_age_all, mean_age_all, sd_mean_age_all, median_age_all,
sd_median_age_all)
```

```
key_figures_age_all


#ggplot(activity_space_age_all, aes(age, y = value, color = variable)) +
#  geom_point(aes(y = mean_space_age, col = "blue")) +
#  geom_line(aes(y = median_space_age, col = "red"))+
#  labs(title = "Distribution of the activity spaces by age, all participants", x = "age of the
participants", y = "Activity Space area [km^2]", color = "") +
#  scale_color_manual(labels = c("mean", "median"), values = c("blue", "red"))
#summary(participants_activity_df)


#####
# distribution if families are filtered, income similar? Influence on activity space if
children
#####
participants_activity_df_age_filtered <- participants_activity_df %>%
  filter(!is.na(age)) %>%
  filter(household_size > 2) %>%
  filter(income <= 3) %>%
  filter(income != 1)
boxplot(ellipse_area ~ age, participants_activity_df_age_filtered, main = 'Activity space
distribution by age, household>2 income <=3',
        ylab = 'Activity space [km^2]', xlab='Age of the participants',
        ylim = c(0, 8500))#, las = 1) +
#theme(axis.title.y = element_text(hjust=-1.5))
mean_activity_space_age <- participants_activity_df_age_filtered %>%
  group_by(age) %>%
  #mutate(space_age = c(ellipse_area)) #%>%
  summarise(mean_space_age = mean(ellipse_area))
median_activity_space_age <- participants_activity_df_age_filtered %>%
  group_by(age) %>%
  #mutate(num_new_activities = sum(new_activities))%>%
  summarise(median_space_age = median(ellipse_area))
activity_space_age <- bind_cols(mean_activity_space_age, median_activity_space_age %>% select(-
age))
activity_space_age

ggplot(activity_space_age, aes(age, y = value, color = variable)) +
  geom_point(aes(y = mean_space_age, col = "blue")) +
  geom_line(aes(y = median_space_age, col = "red"))+
  labs(title = "Activity space distribution by age, household>2 income <=3", x = "age of the
participants", y = "Activity Space area [km^2]", color = "") +
  scale_color_manual(labels = c("mean", "median"), values = c("blue", "red"))



#####
# household_size
#####
participants_activity_df_household_all <- participants_activity_df_all %>%
  filter(!is.na(household_size))
length_household_all <- nrow(participants_activity_df_household_all)
length_household_all

mean_activity_space_household_all <- participants_activity_df_household_all %>%
  group_by(household_size) %>%
  summarise(mean_space_household = mean(ellipse_area))
median_activity_space_household_all <- participants_activity_df_household_all %>%
  group_by(household_size) %>%
  summarise(median_space_household = median(ellipse_area))
std_activity_space_household_all <- participants_activity_df_household_all %>%
  group_by(household_size) %>%
  summarise(std_space_household = sd(ellipse_area))
length_household__seq_all <- participants_activity_df_household_all %>%
  group_by(household_size) %>%
  mutate(rownumberhousehold = row_number())%>%
  group_by(household_size) %>%
  summarise(length_space_household = max(rownumberhousehold))
activity_space_household_all <- bind_cols(mean_activity_space_household_all,
```

```
median_activity_space_household_all %>% select(-household_size),
                                          std_activity_space_household_all %>% select(-
household_size),
                                          length_household__seq_all %>% select(-
household_size))
activity_space_household_all

boxplot(ellipse_area ~ household_size, participants_activity_df_household_all,
        cex.main = 1.2, cex.axis = 1.2, cex.lab = 1.4,
        main = '',
        ylab = 'Activity space [km2]', xlab='household size',
        ylim = c(0, 6000), las = 1)

mean_household_all <- mean(activity_space_household_all$mean_space_household)
median_household_all <- median(activity_space_household_all$median_space_household)
sd_mean_household_all <- sd(activity_space_household_all$mean_space_household)
sd_median_household_all <-  sd(activity_space_household_all$median_space_household)

key_figures_household_all <- c(length_household_all, mean_household_all, sd_mean_household_all,
median_household_all, sd_median_household_all)
key_figures_household_all

#ggplot(activity_space_household_all, aes(household_size, y = value, color = variable)) +
#  geom_point(aes(y = mean_space_household, col = "blue"), size = 3) +
#  geom_point(aes(y = median_space_household, col = "red"), size = 3)+
#  labs(title = "Distribution of activity space area by household size, all participants", x =
"household size", y = "Activity space area [km2]", color = "variable") +
#  scale_color_manual(labels = c("mean", "median"), values = c("blue", "red"))
# regModel__household <- lm(mean_space_household ~ household_size, data =
mean_activity_space_household)
# regModel__household
# abline(regModel__household$coef[1], regModel_mean$coef[2], col = 'red', lwd = 3)
# summary(regModel__householdn)
# model_mean_household <- poly(mean_space ~ age, data = mean_activity_space_household)


#####
# activity space vs income
#####
participants_activity_df_income_all <- participants_activity_df_all %>%
  filter(income != 99)
length_income_all <- nrow(participants_activity_df_income_all)
length_income_all

mean_activity_space_income_all <- participants_activity_df_income_all %>%
  group_by(income) %>%
  summarise(mean_space_income = mean(ellipse_area))
median_activity_space_income_all <- participants_activity_df_income_all %>%
  group_by(income) %>%
  summarise(median_space_income = median(ellipse_area))
std_activity_space_income_all <- participants_activity_df_income_all %>%
  group_by(income) %>%
  summarise(std_space_income = sd(ellipse_area))
length_income__seq_all <- participants_activity_df_income_all %>%
  group_by(income) %>%
  mutate(rownumberincome = row_number())%>%
  group_by(income) %>%
  summarise(length_space_income = max(rownumberincome))
activity_space_income_all <- bind_cols(mean_activity_space_income_all,
median_activity_space_income_all %>% select(-income),
                                       std_activity_space_income_all %>% select(-income),
length_income__seq_all%>% select(-income))
activity_space_income_all

boxplot(ellipse_area ~ income, participants_activity_df_income_all,
        cex.main = 1.2, cex.axis = 1.2, cex.lab = 1.4,
        main = '',
        ylab = 'Activity space [km2]', xlab='income group ',
        ylim = c(0, 6000), las = 1)
```

```
mean_income_all <- mean(activity_space_income_all$mean_space_income)
median_income_all <- median(activity_space_income_all$median_space_income)
sd_mean_income_all <- sd(activity_space_income_all$mean_space_income)
sd_median_income_all <-  sd(activity_space_income_all$median_space_income)

key_figures_income_all <- c(length_income_all, mean_income_all, sd_mean_income_all,
median_income_all, sd_median_income_all)
key_figures_income_all


#plot(activity_space_income$income, activity_space_income$mean_space_income)
#income_label <- c('<4´000', '4´001-8´000', '8´001-12´000', '12´001-16´000', '>16´000')
#ggplot(activity_space_income, aes(income, y = value, color = variable)) +
#  geom_point(aes(y = mean_space_income, col = "blue"), size = 3) +
#  geom_point(aes(y = median_space_income, col = "red"), size = 3)+
#  labs(title = "Distribution of activity space area by income, all participants", x =
"income", y = "Actvitiy space area [km2]", color = "") +
#  scale_color_manual(labels = c("mean", "median"), values = c("blue", "red"))
#summary(participants_activity_df_filtered)

#plot(participants_activity_df_filtered$income, participants_activity_df_filtered$ellipse_area,
ylim = c(0,5000))
#abline(participants_activity_df$age~ participants_activity_df$ellipse_area)
#regModel <- lm(ellipse_area ~ income, data = participants_activity_df)
#regModel
#abline(regModel$coef[1], regModel$coef[2], col = 'red', lwd = 3)
#summary(regModel)



######
# gender
######
participants_activity_df_sex_all <- participants_activity_df_all %>%
  filter(!is.na(sex)) %>%
  mutate(sex = ifelse(sex == 1, "Male", "Female"))
length_sex_all <- nrow(participants_activity_df_sex_all)
length_sex_all

boxplot(ellipse_area ~ sex, participants_activity_df_sex_all,
        cex.main = 1.2, cex.axis = 1.2, cex.lab = 1.4,
        main = 'Activity space distribution by gender, all participants',
        ylab = 'Activity space [km2]', xlab='gender',
        ylim = c(0, 4500), las = 1)

#####
# activity space vs culture/language
#####
participants_activity_df_language_all <- participants_activity_df_all %>%
  filter(!is.na(language))
length_language_all <- nrow(participants_activity_df_language_all)
length_language_all

mean_activity_space_language_all <- participants_activity_df_language_all %>%
  group_by(language) %>%
  summarise(mean_space_language = mean(ellipse_area))
median_activity_space_language_all <- participants_activity_df_language_all %>%
  group_by(language) %>%
  summarise(median_space_language = median(ellipse_area))
sd_activity_space_language_all <- participants_activity_df_language_all %>%
  group_by(language) %>%
  summarise(sd_space_language = sd(ellipse_area))
length_language__seq_all <- participants_activity_df_language_all %>%
  group_by(language) %>%
  mutate(rownumberlanguage = row_number())%>%
  group_by(language) %>%
  summarise(length_space_language = max(rownumberlanguage))
activity_space_language_all <- bind_cols(mean_activity_space_language_all,
```

```
median_activity_space_language_all %>% select(-language),
                                        sd_activity_space_language_all%>% select(-language),
length_language__seq_all%>% select(-language))
activity_space_language_all

boxplot(ellipse_area ~ language, participants_activity_df_language,
        cex.main = 1.2, cex.axis = 1.2, cex.lab = 1.4,
        main = '',
        ylab = 'Activity space [km2]', xlab='language',
        ylim = c(0, 3500), las = 1)

mean_language_all <- mean(activity_space_language_all$mean_space_language)
median_language_all <- median(activity_space_language_all$median_space_language)
sd_mean_language_all <- sd(activity_space_language_all$mean_space_language)
sd_median_language_all <-  sd(activity_space_language_all$median_space_language)

key_figures_language_all <- c(length_language_all, mean_language_all, sd_mean_language_all,
median_language_all, sd_median_language_all)
key_figures_language_all

#plot(activity_space_language$language, activity_space_language$mean_space_language)
#ggplot(activity_space_language, aes(language, y = value, color = variable)) +
#  geom_point(aes(y = mean_space_language, col = "blue"), size = 3) +
#  geom_point(aes(y = median_space_language, col = "red"), size = 3)+
#  labs(title = "Distribution of activity space area by language, control group", x =
"language", y = "Actvitiy space area [km2]", color = "") +
#  scale_color_manual(labels = c("mean", "median"), values = c("blue", "red"))
#summary(participants_activity_df_language)


#####

control <- key_figures_control
nudging <- key_figures_nudging
pricing <- key_figures_pricing
all_participants <- key_figures_all

label <- c('length', 'mean', 'median', 'std. deviation', '25%-Quantile', '75%-Quantile')
key_figures_df_all <- data.frame(label, control, nudging,
                                 pricing, all_participants)
key_figures_df_all
```

B.4 R-Code: Innovation rate


```
# innovation and new places visited
library(dplyr)
library(sf)
library(car)
library(ggmap)
library(ggplot2)
library(lubridate)
library(sp)
library(dbscan)
library(RColorBrewer)

load("/data/mobis/data/mobis_r_workspace.RData")
# mobis_activities_df: activities and accessability of all participants
# mobis_legs_df
# participants: all participants


## use to determine participants for innovation
innovation_activities <- mobis_activities_df %>%
  filter(user_id %in% unique(teilnehmer$participant_ID)) %>%
  filter(phase <= 2) %>%
  filter(in_switzerland == TRUE)
# coordinate structure, change into reference system EPSG:2056 to calculate ellipses
my_activities_sdf_8weeks <- st_sf(my_activities_8weeks)
my_activities_sdf_8weeks$geometry <- st_transform(my_activities_sdf_8weeks$geometry, crs =
2056)
my_activities_sdf_8weeks

## Use for filtered participants
innovation_activities <- mobis_activities_df %>%
  filter(user_id %in% unique(teilnehmer_innovation$participant_ID)) %>%
  filter(phase <= 2) %>%
  filter(in_switzerland == TRUE)

## coordinate structure, change into reference system EPSG:2056 to calculate ellipses
my_activities_sdf_innovation <- st_sf(innovation_activities)
my_activities_sdf_innovation$geometry <- st_transform(my_activities_sdf_innovation$geometry,
crs = 2056)

## get innovation for one person/user; cluster first and let the earliest date for location.
#this will creates cluster id's for eps values of 10,20 and 50 meters (if the geometries are in
EPSG:2056 or similar: you can check this with st_crs(activities_df))
#an eps of 10 means that points within in 10 units (in our case meters) will be clustered into
the same group.
#min points means that a cluster must have at least 3 point in it - you can set this to 1 to
have a cluster id for every activity, but that isn't really done.
#activities with a cluster value of 0 are outliers with no id - ie one of stops - maybe you can
flag these as a new activity if the duration of the stay was of a certain length.
#activities with the same id are essentially performed at the same location.
db2 <- function(x) {
  activity_coordinates <- x  %>% st_coordinates()
  #cluster_10 = dbscan::dbscan(activity_coordinates, eps = 10, minPts = 3)$cluster
  #cluster_20 = dbscan::dbscan(activity_coordinates, eps = 20, minPts = 3)$cluster
  cluster_50 = dbscan::dbscan(activity_coordinates, eps = 50, minPts = 3)$cluster
  #return (data.frame(cluster_10m=cluster_10,cluster_20m=cluster_20, cluster_50m=cluster_50))
  return (data.frame(cluster_50m=cluster_50))
}

#####
#innovation for one particular user
#####
#select a user you want to analyse
user <- 'MCNDV'
# define the data you want to plot
my_activities_8weeks_user <- my_activities_sdf_8weeks %>%
  filter(user_id == user)

# get matrix with clustered activities
# use the dbscan algorithm to cluster activites
```

```
activity_cluster_df_user <- my_activities_8weeks_user %>%
  group_by(user_id) %>%
  group_modify(~db2(.x)) %>% ungroup()

#join this back to the activties, dont duplicate the user_id column
activities_df_with_clusters_user <- bind_cols(my_activities_8weeks_user,
activity_cluster_df_user %>%
                                                select(-user_id))

one_person_activities_user = activities_df_with_clusters_user %>%
  # filter(user_id == 'AACZJ') %>% #only take this person
  group_by(cluster_50m) %>% #group by cluster
  filter(cluster_50m != 0) %>% # don't consider outliers as a cluster group
  mutate(cluster_size=n()) %>% #get the size of each cluster (this will then add it as a new
column) - kinda like summarize, but it returns all the original data rows
  ungroup() %>% #remove the grouping
  mutate(cluster_rank=dense_rank(desc(cluster_size))) %>% #rank the clusters by size - largest
first
  filter(cluster_rank <= 5) #take the top 5 clusters
one_person_activities_user

#get the map for this area:
bbox <- one_person_activities_user %>%
  st_bbox() %>%
  st_as_sfc() %>%
  st_buffer(10000) %>%
  st_transform(4326) %>%
  st_bbox() %>%
  setNames(c('left', 'bottom', 'right', 'top'))

# use the bounding box to get a stamen map
map <- get_map(bbox)

#if you plot the activities for one participant, and color by cluster id, you can see how this
works
# you can see that cluster 4 is hidden behind 5.
ggmap(map) +
  #   geom_sf(data=kantons, color='black') +
  geom_sf_label(data=st_transform(one_person_activities_user, 4326),
                mapping=aes(color = as.factor(cluster_50m), label=cluster_50m, size =
cluster_size),
                inherit.aes = FALSE) +
  labs(title='Top 5 activity locations for user MCNDV', subtitle = '(50m eps dbscan clusters,
min. points for a cluster = 3)',
       color = 'number of cluster', size = 'cluster size')


#####
#innovation for all user
#####

#work with a subset for the moment - but change this if you need to.
my_activites_df_innovation <- my_activities_sdf_innovation %>% arrange(user_id, started_at) #
%>% slice(1:10000)

#use the dbscan algorithm above to cluster activites
activity_cluster_df <- my_activites_df_innovation %>%
  group_by(user_id) %>%
  group_modify(~db2(.x)) %>% ungroup()

#join this back to the activties, dont duplicate the user_id column
activities_df_with_clusters <- bind_cols(my_activites_df_innovation, activity_cluster_df %>%
select(-user_id))

##### This block gives id's to the outliers
non_outlier_ativities <- activities_df_with_clusters %>% filter(cluster_50m > 0) %>%
mutate(cluster_50m_no_outliers = cluster_50m)
#give outliers a cluster_id as well
outlier_activities <- activities_df_with_clusters %>%
```

```
  group_by(user_id) %>%
  mutate(max_cluster_id = max(cluster_50m)) %>%
  filter(cluster_50m == 0) %>% # don't consider outliers as a cluster group
  arrange(started_at) %>%
  mutate(cluster_50m_no_outliers = max_cluster_id + row_number())%>%
  select(-max_cluster_id)

re_clustered_activities_df <- rbind(non_outlier_ativities, outlier_activities)


### this block sets to true if the activity location wasn't seen before.
re_clustered_activities_df1 <- re_clustered_activities_df %>%
  arrange(user_id, started_at) %>%
  group_by(user_id) %>%
  filter(cluster_50m > 0) %>% #set this if you don't want to include the outliers (which have a
cluster_50m id of 0)
  mutate(cluster_chronological_id = match(cluster_50m_no_outliers,
unique(cluster_50m_no_outliers))) %>% #this line calculates new ids in cronological order for
each person.
  group_by(user_id, day_g=floor_date(started_at, 'day')) %>%  ### set day_g to what you want to
indicate a new location by.
  mutate(max_crono_id_day = max(cluster_chronological_id))

#get the max chronological id for the previous day
previous_max_crono_id_days_s <- re_clustered_activities_df1 %>%
  st_drop_geometry() %>%
  group_by(user_id, day_g) %>%
  summarize(max_crono_id_day=max(max_crono_id_day)) %>%
  group_by(user_id) %>%
  mutate(prev_max_crono_id_day = lag(max_crono_id_day, default=0))

#Check each day to see if the chronological ids are greater than the max from the previous day,
if so, it is a new location.
re_clustered_activities_df_with_new_activity_indicator <- re_clustered_activities_df1 %>%
left_join(previous_max_crono_id_days_s) %>%
  group_by(user_id) %>%
  mutate(is_new_activity_location_on_day = cluster_chronological_id > prev_max_crono_id_day)
%>%
  select(-day_g, -max_crono_id_day, -prev_max_crono_id_day)

re_clustered_activities_df_with_new_activity_indicator

a <- re_clustered_activities_df_with_new_activity_indicator %>%
  st_drop_geometry() %>% # we dont need to work with the geometries here, makes it much faster
  group_by(user_id, day=floor_date(started_at, 'day')) %>%
  summarize(new_activities = sum(is_new_activity_location_on_day)) #%>%
  #filter(user_id == 'MCNDV')

### grouped by ID and date, all
activities_df_with_clusters %>%
  group_by(user_id, floor_date(started_at, 'day')) %>%
  summarise(daily_cluster_ids = list(unique(cluster_50m)))


### get the days numbered in order to compare day 1 for all users no matter what date it was
a_num_date <- a %>%
  group_by(user_id)%>%
  mutate(n_day = row_number())%>%
  ungroup()#%>%
  #slice(1:560)

innovation_after_4weeks <- a_num_date %>%
  filter(n_day > 28)

#####
# determine which participants need to take into consideration, use the total data set and
filter after you got the ones longer than 54 days
#####
obs_time <- a_num_date %>%
```

```
  group_by(user_id) %>%
  summarise(num_obs_day = max(n_day))

num_teilnehmer_inno <- obs_time %>%
  filter(num_obs_day >= 52)
setwd('~/Test data/Data') ##Set to where you want your file
write_csv(num_teilnehmer_inno, 'teilnehmer_innovation.csv')
#####

### line plot for all new activities, these are not in order for days, rather in chronological
sequence
ggplot(a) +
  geom_line(aes(x=day, y=new_activities))+
  labs(title = "New activities, control group", x = "Observation period", y = "new activity
locations", color = "")

### line plot of new activities by observation day, no matter what date the activity occured
#ggplot(a_num_date, aes(x=n_day, y=new_activities)) +
#  geom_line(shape=1) +    # Use hollow circles
#  geom_smooth(method = loess)   # Add linear regression line
#  (by default includes 95% confidence region)



#####
# plot point to get the parameters for the ggplot
#####
plot(a_num_date$n_day, a_num_date$new_activities, xlab = 'Day of observation', ylab = 'New
activities', main = 'Number of new activities by day of observation')
regModel <- lm(new_activities ~ n_day, data = a_num_date)
regModel
abline(regModel$coef[1], regModel$coef[2], col = 'red', lwd = 3)
regModel_day_new_activities <- summary(regModel)
regModel_day_new_activities
regModel$coef[1]

### plot in day order no matter of date
ggplot(a_num_date) +
  geom_line(aes(x=n_day, y=new_activities))+
  #stat_summary(fun.data= mean_cl_normal) +
  #geom_smooth(method='lm')+
  geom_abline(aes(intercept = regModel$coef[1], slope = regModel$coef[2]), col = 'red', size =
1.5)+
  #scale_color_brewer(palette="Spectral")+
  labs(title = "New activities for total innovation", x = "Observation period [days]", y = "new
activity locations", color = "")+
  theme(plot.title = element_text(color="black", size=20),
        axis.title.x = element_text(color="black", size=16),
        axis.title.y = element_text(color="black", size=16),
        axis.text = element_text(size = 14))

#####
#plot in day order no matter of date, after 4 weeks passed
#####
plot(innovation_after_4weeks$n_day, innovation_after_4weeks$new_activities, xlab = 'Day of
observation', ylab = 'New activities', main = 'Number of new activities by day of observation')
regModel_4weeks <- lm(new_activities ~ n_day, data = innovation_after_4weeks)
regModel_4weeks
abline(regModel_4weeks$coef[1], regModel_4weeks$coef[2], col = 'red', lwd = 3)
regModel_day_new_activities_4weeks <- summary(regModel_4weeks)
regModel_day_new_activities_4weeks
regModel_4weeks$coef[1]

ggplot(innovation_after_4weeks) +
  geom_line(aes(x=n_day, y=new_activities))+
  #stat_summary(fun.data= mean_cl_normal) +
  #geom_smooth(method='lm')+
  geom_abline(aes(intercept = regModel_4weeks$coef[1], slope = regModel_4weeks$coef[2]), col =
'red', size = 1.2)+
```

```
  #scale_color_brewer(palette="Spectral")+
  labs(title = "New activities for innovation after four weeks", x = "Observation period
[days]", y = "new activity locations", color = "")+
  theme(plot.title = element_text(color="black", size=20),
        axis.title.x = element_text(color="black", size=16),
        axis.title.y = element_text(color="black", size=16),
        axis.text = element_text(size = 14))

#### group for id and get a rate for each user
a_num_date_innovation <- a_num_date %>%
  group_by(user_id)%>%
  #mutate(inno_rate = mean(a_num_date$new_activities))%>%
  summarise(rate = mean(new_activities))
a_num_date_innovation_4weeks <- innovation_after_4weeks %>%
  group_by(user_id)%>%
  #mutate(inno_rate = mean(a_num_date$new_activities))%>%
  summarise(rate_4week = mean(new_activities))

innovation_rate_filter <- a_num_date_innovation %>%
  filter(user_id %in% unique(a_num_date_innovation_4weeks$user_id))
innovation_rate_compare <- bind_cols(innovation_rate_filter, a_num_date_innovation_4weeks %>%
select(-user_id))

innovation_rate <- mean(innovation_rate_compare$rate)
innovation_rate_4weeks <- mean(innovation_rate_compare$rate_4week)
innovation_rate
innovation_rate_4weeks


#####
# add participants data to rate
#####
innovation_participants_df_filtered <- teilnehmer_innovation %>%
  filter(participant_ID %in% unique(innovation_rate_compare$user_id))
innovation_participants_df <- bind_cols(innovation_participants_df_filtered,
innovation_rate_compare)

### key figures for the participant_innovation data
label <- c('length', 'mean', 'median', 'std. deviation', '25%-Quantile', '75%-Quantile')
length_inno <- nrow(innovation_participants_df)
mean_inno <- mean(innovation_participants_df$rate)
median_inno <- median(innovation_participants_df$rate)
std_inno <- sd(innovation_participants_df$rate)
quantile1_inno <- quantile(innovation_participants_df$rate, 0.25)
quantile2_inno <- quantile(innovation_participants_df$rate, 0.75)
figures_innovation_total <- c(length_inno, mean_inno, median_inno, std_inno, quantile1_inno,
quantile2_inno)

length_inno_4week <- nrow(innovation_participants_df)
mean_inno_4week <- mean(innovation_participants_df$rate_4week)
median_inno_4week <- median(innovation_participants_df$rate_4week)
std_inno_4week <- sd(innovation_participants_df$rate_4week)
quantile1_inno_4week <- quantile(innovation_participants_df$rate_4week, 0.25)
quantile2_inno_4week <- quantile(innovation_participants_df$rate_4week, 0.75)
figures_innovation_4week <- c(length_inno_4week, mean_inno_4week, median_inno_4week,
std_inno_4week, quantile1_inno_4week, quantile2_inno_4week)

figures_innovation <- data_frame(label, figures_innovation_total, figures_innovation_4week)
figures_innovation

innovation_participants_df$AgeGroup <- cut(innovation_participants_df$age, breaks = c(seq(10,
65, by = 10), Inf))
boxplot(rate ~ AgeGroup, innovation_participants_df,
        cex.main = 1.6, cex.axis = 1.3, cex.lab = 1.3,
        main = 'Innovation rate over the whole observation time',
        ylab = 'Innovation rate [new activities/day]', xlab='All Participants',
        ylim = c(0.2, 2.6), las = 1)

#####
```

```
# innovation rate after 4 weeks
#####
innovation_participants_df_age <- innovation_participants_df %>%
  filter(!is.na(age))

boxplot(rate_4week ~ age, innovation_participants_df_age,
        cex.main = 1.4, cex.axis = 1.3, cex.lab = 1.3,
        main = 'Innovation rate after four weeks of observation',
        ylab = 'Innovation rate [new activities/day]', xlab='Age of the participants',
        ylim = c(0.2, 2.6), las = 1)
innovation_participants_df_age$AgeGroup <- cut(innovation_participants_df_age$age, breaks =
c(seq(10, 65, by = 10), Inf))
boxplot(rate_4week ~ AgeGroup, innovation_participants_df_age,
        cex.main = 1.6, cex.axis = 1.8, cex.lab = 1.8,
        main = '',
        ylab = 'Innovation rate [new activities/day]', xlab='Age of the participants',
        ylim = c(0.2, 2.6), las = 1)


mean_innovation_age <- innovation_participants_df_age %>%
  group_by(AgeGroup)  %>%
  #mutate(space_age = c(ellipse_area)) #%>%
  summarise(mean_inno_age = mean(rate_4week))
median_innovation_age <- innovation_participants_df_age %>%
  group_by(AgeGroup) %>%
  #mutate(num_new_activities = sum(new_activities))%>%
  summarise(median_inno_age = median(rate_4week))
sd_innovation_age <- innovation_participants_df_age %>%
  group_by(AgeGroup) %>%
  #mutate(num_new_activities = sum(new_activities))%>%
  summarise(sd_inno_age = sd(rate_4week))
length_innovation_AgeGroup <- innovation_participants_df_age %>%
  group_by(AgeGroup) %>%
  mutate(rownumberAge = row_number())%>%
  group_by(AgeGroup) %>%
  summarise(length_space_age = max(rownumberAge))
innovation_age <- bind_cols(mean_innovation_age, median_innovation_age %>% select(-AgeGroup),
sd_innovation_age %>% select(-AgeGroup),
                            length_innovation_AgeGroup %>% select(-AgeGroup))
innovation_age

#####
# innovation vs. income
#####
innovation_participants_df_income <- innovation_participants_df%>%
  filter(income != 99)

boxplot(rate_4week ~ income, innovation_participants_df_income,
        cex.main = 1.6, cex.axis = 1.8, cex.lab = 1.8,
        main = '',
        ylab = 'Innovation rate [new activities/day]', xlab='Income group',
        ylim = c(0.2, 2.6), las = 1)

mean_innovation_income <- innovation_participants_df_income %>%
  group_by(income)  %>%
  #mutate(space_income = c(ellipse_area)) #%>%
  summarise(mean_inno_income = mean(rate_4week))
median_innovation_income <- innovation_participants_df_income %>%
  group_by(income) %>%
  #mutate(num_new_activities = sum(new_activities))%>%
  summarise(median_inno_income = median(rate_4week))
sd_innovation_income <- innovation_participants_df_income %>%
  group_by(income) %>%
  #mutate(num_new_activities = sum(new_activities))%>%
  summarise(sd_inno_income = sd(rate_4week))
length_innovation_income <- innovation_participants_df_income %>%
  group_by(income) %>%
  mutate(rownumberincome = row_number())%>%
  group_by(income) %>%
```

```
  summarise(length_space_income = max(rownumberincome))
innovation_income <- bind_cols(mean_innovation_income, median_innovation_income %>% select(-
income), sd_innovation_income %>% select(-income),
                              length_innovation_income %>% select(-income))
innovation_income

#####
# innovation vs. household
#####
innovation_participants_df_household <- innovation_participants_df%>%
  filter(!is.na(household_size))

boxplot(rate_4week ~ household_size, innovation_participants_df_household,
        cex.main = 1.6, cex.axis = 1.8, cex.lab = 1.8,
        main = '',
        ylab = 'Innovation rate [new activities/day]', xlab='household size',
        ylim = c(0.2, 2.6), las = 1)

mean_innovation_household <- innovation_participants_df_household %>%
  group_by(household_size)  %>%
  #mutate(space_household = c(ellipse_area)) #%>%
  summarise(mean_inno_household = mean(rate_4week))
median_innovation_household <- innovation_participants_df_household %>%
  group_by(household_size) %>%
  #mutate(num_new_activities = sum(new_activities))%>%
  summarise(median_inno_household = median(rate_4week))
sd_innovation_household <- innovation_participants_df_household %>%
  group_by(household_size) %>%
  #mutate(num_new_activities = sum(new_activities))%>%
  summarise(sd_inno_household = sd(rate_4week))
length_innovation_household <- innovation_participants_df_household %>%
  group_by(household_size) %>%
  mutate(rownumberhousehold = row_number())%>%
  group_by(household_size) %>%
  summarise(length_space_household = max(rownumberhousehold))
innovation_household <- bind_cols(mean_innovation_household, median_innovation_household %>%
select(-household_size),
                                  sd_innovation_household %>% select(-household_size),
length_innovation_household %>% select(-household_size))
innovation_household

###############################################################################
#gender
#####
innovation_participants_df_sex <- innovation_participants_df %>%
  mutate(sex = ifelse(sex == 1, "Male", "Female")) %>%
  filter(!is.na(sex))

boxplot(rate_4week ~ sex, innovation_participants_df_sex,
        cex.main = 1.6, cex.axis = 1.3, cex.lab = 1.3,
        main = 'Innovation rate after four weeks of observation',
        ylab = 'Innovation rate [new activities/day]', xlab='gender',
        ylim = c(0.2, 2.6), las = 1)

mean_innovation_gender <- innovation_participants_df_sex %>%
  group_by(sex)  %>%
  #mutate(space_gender = c(ellipse_area)) #%>%
  summarise(mean_inno_gender = mean(rate_4week))
median_innovation_gender <- innovation_participants_df_sex %>%
  group_by(sex) %>%
  #mutate(num_new_activities = sum(new_activities))%>%
  summarise(median_inno_gender = median(rate_4week))
sd_innovation_gender <- innovation_participants_df_sex %>%
  group_by(sex) %>%
  #mutate(num_new_activities = sum(new_activities))%>%
  summarise(sd_inno_gender = sd(rate_4week))
innovation_gender <- bind_cols(mean_innovation_gender, median_innovation_gender %>% select(-
sex), sd_innovation_gender %>%select((-sex)))
innovation_gender
```

```
######
# culture
#####
innovation_participants_df_language <- innovation_participants_df %>%
  filter(!is.na(language))
length_language_inno <- nrow(innovation_participants_df_language)

boxplot(rate_4week ~ language, innovation_participants_df_language,
        cex.main = 1.6, cex.axis = 1.8, cex.lab = 1.8,
        main = '',
        ylab = 'Innovation rate [new activities/day]', xlab='language',
        ylim = c(0.2, 2.6), las = 1)

mean_innovation_language <- innovation_participants_df_language %>%
  group_by(language) %>%
  #mutate(num_new_activities = sum(new_activities))%>%
  summarise(mean_innovation_language = mean(rate_4week))
median_innovation_language <- innovation_participants_df_language %>%
  group_by(language) %>%
  #mutate(num_new_activities = sum(new_activities))%>%
  summarise(median_innovation_language = median(rate_4week))
sd_innovation_language <- innovation_participants_df_language %>%
  group_by(language) %>%
  #mutate(num_new_activities = sum(new_activities))%>%
  summarise(sd_inno_language = sd(rate_4week))
length_innovation_language <- innovation_participants_df_language %>%
  group_by(language) %>%
  mutate(rownumberlanguage = row_number())%>%
  group_by(language) %>%
  summarise(length_space_language = max(rownumberlanguage))
innovation_language <- bind_cols(mean_innovation_language, median_innovation_language %>%
select(-language),
                                 sd_innovation_language %>% select(-language),
length_innovation_language %>% select(-language))
innovation_language


#####
# Participants for innovation, plots
#####
num_teilnehmer_inno

teilnehmer_innovation <- total_participants %>%
  filter(participant_ID %in% unique(num_teilnehmer_inno$user_id))

# AGE
#####
hist(teilnehmer_innovation$age, breaks = 25, col = 'pink',
     cex.main = 1.5, cex.axis = 1.4, cex.lab = 1.4,
     main = 'Age distribution for the innovation rate',
     xlab = 'Age of the participants', ylab = 'Number of participants', las = 1)

#####
# HOUSEHOLD SIZE
#####
hist(teilnehmer_innovation$household_size)
hist(teilnehmer_innovation$household_size, breaks = 25, col = 'pink',
     cex.main = 1.5, cex.axis = 1.4, cex.lab = 1.4,
     main = 'household size distribution for the innovation rate',
     xlab = 'Household size', ylab = 'Number of participants', las = 1)

#####
# Income
#####
hist(teilnehmer_innovation$income, breaks = 500, col = 'pink',
     cex.main = 1.5, cex.axis = 1.4, cex.lab = 1.4,
     main = 'Income distribution for the innovation rate',
     xlab = 'Income group', ylab = 'Number of participants', las = 1,
```

```
     xlim = c(1,5))

#####
## Output in a csv file
innovation_rate_results <- bind_cols(innovation_participants_df %>%select(-total, -user_id),
obs_time %>% select(-user_id))
setwd("~/mobis_analysis/r/eth_activity_spaces") ##Set to where you want your file
write_csv(innovation_rate_results, 'innovation_rate_results.csv')
```

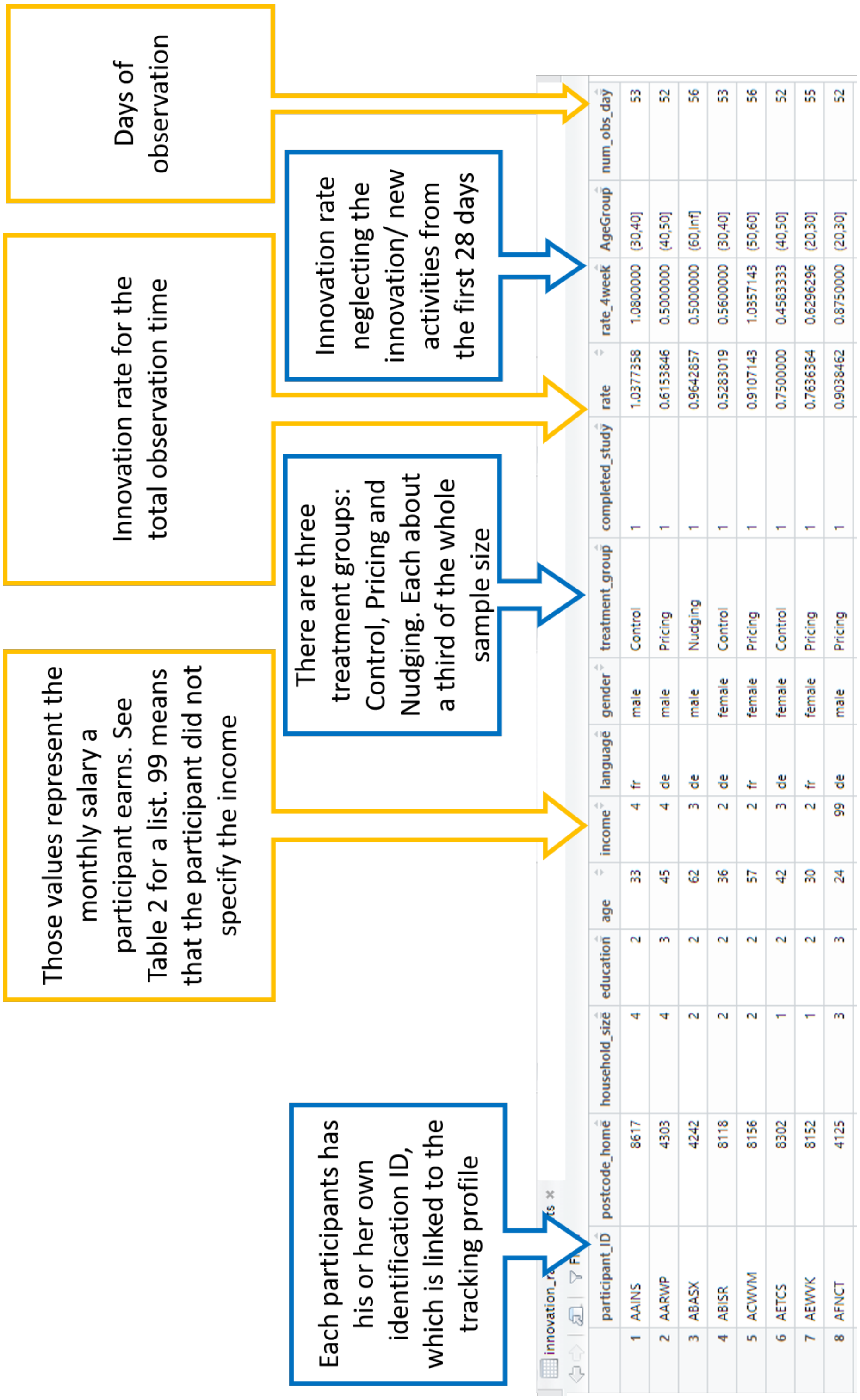
B.5 R-Code: Key Figures Table


```
##### Tables

# Key figures age
mean_age_group <- mean(activity_space_age_group$mean_space_age)
median_age_group <- median(activity_space_age_group$median_space_age)
sd_mean_age_group <- sd(activity_space_age_group$mean_space_age)
sd_median_age_group <-  sd(activity_space_age_group$median_space_age)

key_figures_age_group_control <- c(length_age_group, mean_age_group, sd_mean_age_group,
median_age_group, sd_median_age_group)
key_figures_age_group_control

key_figures_age_group_nudging <- c(length_age_group, mean_age_group, sd_mean_age_group,
median_age_group, sd_median_age_group)
key_figures_age_group_nudging

key_figures_age_group_pricing <- c(length_age_group, mean_age_group, sd_mean_age_group,
median_age_group, sd_median_age_group)
key_figures_age_group_pricing

key_figures_age_group_innovation <- c(length_age_group, mean_age_group, sd_mean_age_group,
median_age_group, sd_median_age_group)
key_figures_age_group_innovation


control <- key_figures_age_group_control
nudging <- key_figures_age_group_nudging
pricing <- key_figures_age_group_pricing
innovation <- key_figures_age_group_innovation
all_participants <- key_figures_age_all

label <- c('length', 'mean', 'std. deviation mean', 'median', 'std. deviation median')
key_figures_df_age <- data.frame(label, control, nudging,
                                 pricing, innovation, all_participants)
key_figures_df_age



# Household size
control <- key_figures_household_group_control
nudging <- key_figures_household_group_nudging
pricing <- key_figures_household_group_pricing
innovation <- key_figures_household_group_innovation
all_participants <- key_figures_household_all

label <- c('Household size', 'mean', 'std. deviation mean', 'median', 'std. deviation median')
key_figures_df_household <- data.frame(label, control, nudging,
                                 pricing, innovation,  all_participants)
key_figures_df_household



# income
control <- key_figures_income_group_control
nudging <- key_figures_income_group_nudging
pricing <- key_figures_income_group_pricing
innovation <- key_figures_income_group_innovation
all_participants <- key_figures_income_all

label <- c('INCOME', 'mean', 'std. deviation mean', 'median', 'std. deviation median')
key_figures_df_income <- data.frame(label, control, nudging,
                                 pricing, innovation,  all_participants)
key_figures_df_income



# summary of all date, not spited into groups
control <- key_figures_control
nudging <- key_figures_nudging
```

```
pricing <- key_figures_pricing
innovation <-  key_figures_innovation
all_participants <- key_figures_all

label <- c('length', 'mean', 'median', 'std. deviation', '25%-Quantile', '75%-Quantile')
key_figures_df_all <- data.frame(label, control, nudging,
                                 pricing,innovation, all_participants)
key_figures_df_all
```

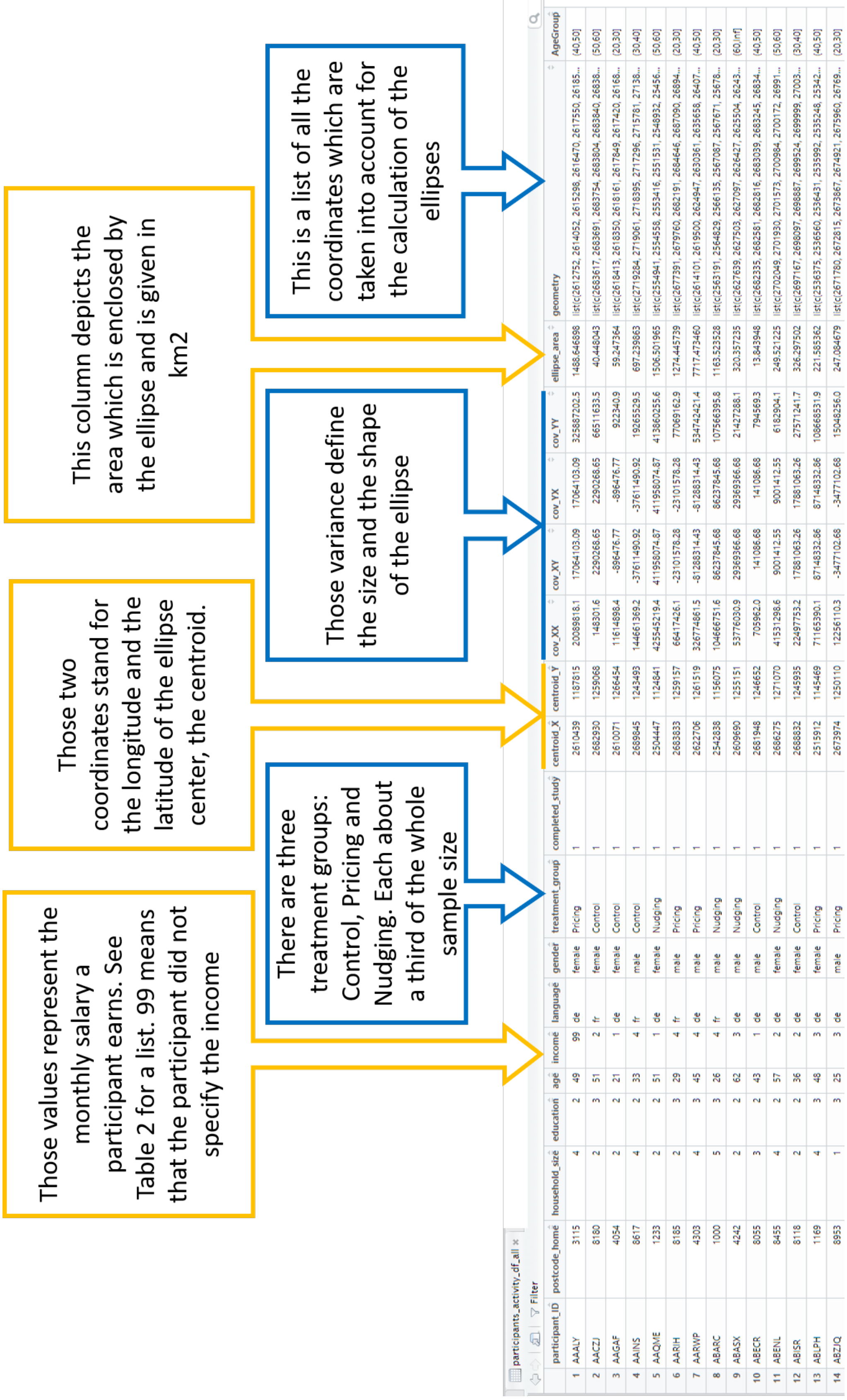
C Documentation Results

C.1 Documentation Results Innovation Rate



	participant_ID	postcode_home	household_size	education	age	income	language	gender	treatment_group	completed_study	rate	rate_4week	AgeGroup	num_obs_day
1	AAINS	8617	4	2	33	4	fr	male	Control	1	1.0377358	1.0800000	(30,40]	53
2	AARWP	4303	4	3	45	4	de	male	Pricing	1	0.6153846	0.5000000	(40,50]	52
3	ABASX	4242	2	2	62	3	de	male	Nudging	1	0.9642857	0.5000000	(60,Inf]	56
4	ABISR	8118	2	2	36	2	de	female	Control	1	0.5283019	0.5600000	(30,40]	53
5	ACWWM	8156	2	2	57	2	fr	female	Pricing	1	0.9107143	1.0357143	(50,60]	56
6	AETCS	8302	1	2	42	3	de	female	Control	1	0.7500000	0.4583333	(40,50]	52
7	AEWVK	8152	1	2	30	2	fr	female	Pricing	1	0.7636364	0.6296296	(20,30]	55
8	AFNCT	4125	3	3	24	99	de	male	Pricing	1	0.9038462	0.8750000	(20,30]	52

C.2 Documentation Results Activity Space



	participant_ID	postcode_home	household_size	education	age	income	language	gender	treatment_group	completed_study	centroid_X	centroid_Y	cov_XX	cov_XY	cov_YX	cov_YY	ellipse_area	geometry	AgeGroup
1	AAALY	3115	4	2	49	99	de	female	Pricing	1	2610439	1187815	20089818.1	17064103.09	17064103.09	325887202.5	1488.646898	list(c(2612752, 2614052, 2615298, 2616470, 2617550, 26185...	(40,50]
2	AACZJ	8180	2	3	51	2	fr	female	Control	1	2682930	1259068	148301.6	2290268.65	2290268.65	66511633.5	40.448043	list(c(2683617, 2683691, 2683754, 2683804, 2683840, 26838...	(50,60]
3	AAGAF	4054	2	2	21	1	de	female	Control	1	2610071	1266454	11614898.4	-896476.77	-896476.77	922340.9	59.247364	list(c(2618413, 2618350, 2618161, 2617849, 2617420, 26168...	(20,30]
4	AAINS	8617	4	2	33	4	fr	male	Control	1	2689845	1243493	144661369.2	-37611490.92	-37611490.92	19265529.5	697.239863	list(c(2719284, 2719061, 2718395, 2717296, 2715781, 27138...	(30,40]
5	AAQME	1233	2	2	51	1	de	female	Nudging	1	2504447	1124841	425545219.4	411958074.87	411958074.87	413860255.6	1506.501965	list(c(2554941, 2554558, 2553416, 2551531, 2548932, 25456...	(50,60]
6	AARIH	8185	2	3	29	4	fr	male	Pricing	1	2683833	1259157	66417426.1	-23101578.28	-23101578.28	77069162.9	1274.445739	list(c(2677391, 2679760, 2682191, 2684646, 2687090, 26894...	(20,30]
7	AARWP	4303	4	3	45	4	de	male	Pricing	1	2622706	1261519	326774861.5	-81288314.43	-81288314.43	534742421.4	7717.473460	list(c(2614101, 2619500, 2624947, 2630361, 2635658, 26407...	(40,50]
8	ABARC	1000	5	3	26	4	fr	male	Nudging	1	2542838	1156075	104666751.6	86237845.68	86237845.68	107566395.8	1163.523528	list(c(2563191, 2564829, 2566135, 2567087, 2567671, 25678...	(20,30]
9	ABASX	4242	2	2	62	3	de	male	Nudging	1	2609690	1255151	53776030.9	29369366.68	29369366.68	21427288.1	320.357235	list(c(2627639, 2627503, 2627097, 2626427, 2625504, 26243...	(60,Inf]
10	ABECR	8055	3	2	43	1	de	male	Control	1	2681948	1246652	705962.0	141086.68	141086.68	794569.3	13.843948	list(c(2682335, 2682581, 2682816, 2683039, 2683245, 26834...	(40,50]
11	ABENL	8455	4	2	57	2	de	female	Nudging	1	2686275	1271070	41531298.6	9001412.55	9001412.55	6182904.1	249.521225	list(c(2702049, 2701930, 2701573, 2700984, 2700172, 26991...	(50,60]
12	ABISR	8118	2	2	36	2	de	female	Control	1	2688832	1245935	22497753.2	17881063.26	17881063.26	27571241.7	326.297502	list(c(2697167, 2698097, 2698887, 2699524, 2699999, 27003...	(30,40]
13	ABLPH	1169	4	3	48	3	de	female	Pricing	1	2515912	1145469	71165390.1	87148332.86	87148332.86	108668531.9	221.585362	list(c(2536375, 2536560, 2536431, 2535992, 2535248, 25342...	(40,50]
14	ABZJQ	8953	1	3	25	3	de	male	Pricing	1	2673974	1250110	12256110.3	-3477102.68	-3477102.68	15048256.0	247.084679	list(c(2671780, 2672815, 2673867, 2674921, 2675960, 26769...	(20,30]