University of Zurich[UZH]

The relationship between visiting places and the health of the older adults: through spatial trajectory data

GEO 511 Master's Thesis

Author
Reetta Välimäki
18-756-544

Supervised by
Dr. Cheng Fu

Faculty representative
Prof. Dr. Robert Weibel

30.09.2020
Department of Geography, University of Zurich

GEO 511 Master's thesis Reetta Välimäki
09/10/2020 18-756-544

The relationship between visiting places and the health of the older adults: through spatial trajectory data

Master's thesis

Reetta Välimäki

Advisors:
Dr. Cheng Fu
Prof. Dr. Robert Weibel

Department of Geography
University of Zurich
Zurich Switzerland
reetta.vaelimaeki@uzh.ch

Table of contents

List of Tables

List of Figures

KEYWORDS
older adults, mobility, health, spatial activities, place modeling, semantic enrichment

1 Introduction

Our environments are increasingly developing as an increasing number of people in the world are living in cities. The places that are incorporated into human life are thus changing drastically. There is also an increasingly aging population in the developed world and cities are being built largely with a working population in mind (Handler, 2014). The relationship between the visitation of places and how they are associated with health is not yet been fully understood.

Spatial activities have been shown to have multiple effects on health and well-being, especially for older adults. Some evidence shows that out-of-home mobility is important for the well-being of the older population and that mobility contributes to overall well-being by delaying the onset of disabilities, promoting healthy aging, and postponing frailty (Mollenkopf et al., 2011, 2004). The number of visited places has also been associated with social network size and place visitation patterns with different personality types (Alessandretti et al., 2018). Isaacson et al. (2017) found a complex relationship between spatial activity and well-being that needs to be further examined to enhance the understanding of which spatial activities can benefit the aging population.

Spatial activities are closely associated with places, as places describe the meanings attached to locations (Tuan, 1975). One type of place may be associated with some specific spatial activity. More research is needed regarding the types of places that lead to increased well-being in the older population, which would lead to a greater understanding of the aging population mobility and how it can be improved. Understanding of the types of places that affect health can also contribute towards a better understanding of important activities that would allow services to be provided according to the needs of the aging population.

In previous research, spatial activities in relation to health are often explored through measures such as the shape of the spatial activity, distance to home, and the duration out of home (Fillekes et al., 2019). These are simple physical measures without much semantic context. Exploring the relationship between health and place visitation may provide a more comprehensive understanding. Some studies use travel diaries or questionnaires to compare the relationship between place visitation and health, e.g., van den Berg et al. (2019), but this can be very limited and is subject to errors. Place visitations have been modeled successfully in previous studies using GPS devices, for example, to model tourism behavior (McKercher et al., 2012; Zheng et al., 2017)

In this study, the relationship between place visitation and aging people's health is explored. So far, time spending and the effects on health using GPS trajectories have been insufficiently studied, and the results are inconclusive. We aimed to investigate this relationship by aggregating visited places based on their semantics derived from the GPS trajectories. The visitation frequencies of the individual and aggregated places were then compared to health indicators.

We aimed to find a relationship between the older adults' place visitation and health. We extracted the stop episodes with their semantics and used novel aggregation methods to cluster them into lower granularities. We then compared the relationships between the health indicators and visitation of individual and aggregated places. This allowed the examination of whether there is a relationship between health and place visitations as well as which methods can best capture this relationship.

The thesis is part of the MOASIS project, which includes 159 over 65-year-olds participants who carried GPS loggers for approximately 30 days and were surveyed for mental and physical health. The data were collected according to the University of Zurich's Ethics Policy. The participants gave written consent for the usage of their data.

1.1 Structure of the thesis

The thesis starts with a state-of-the-art literature review (Chapter 2) concerning the spatial activities of the aging population in relation to health. Chapter 3 introduces the goal of the research. Chapter 4 gives more details about the structure of the data used in this study and how they were collected. The methodology section (Chapter 5) first explains how the GPS waypoints were processed and enriched by places. The following section describes how a variety of place clustering methods were applied to aggregate places into more general place concepts. These will then be followed by results of the multivariate regressions between place visitation and health (Chapter 6). Chapter 7 introduces the limitations of this study and how they might have impacted the results. Chapter 8 concludes the main findings of the thesis and proposes possible future studies.

2 Related work

2.1 Spatial activities and the relation to health in the older population

Lemon et al. (1972) defined an activity as "any regularized or patterned action or pursuit which is regarded beyond routine physical or personal maintenance." They also divided it into informal, formal, and solitary activities, where informal includes activities with friends, relatives, and neighbors and formal includes voluntary organizations. Health can be defined as "A state of complete physical, mental, and social wellbeing not merely the absence of disease or infirmity" (WHO, 2019, 1948). This is a holistic approach to health that incorporates the quality of life and general wellbeing into the definition. In gerontology, health is most often considered the primary indicator of wellbeing (Enam et al., 2018).

The activity theory of aging was formally formulated by Lemon et al. in 1972. It proposes that there is a positive relationship between activity levels and life satisfaction in the aging population. The activity theory has received mixed results. Lemon et al. (1972) found a positive relation between informal activities and life satisfaction, but no relationship between formal and solitary activities compared to life satisfaction was found. Longino and Kart (1982) found a positive correlation between informal activities and life satisfaction, whereas solitary activities had no effect, and formal activities had a negative effect.

Enam et al. (2018) discovered a negative relation between the aging population's time expenditure in activity participation and self-perceived health and wellbeing. Isaacson et al. (2017) found a relationship between perceived physical functioning, out-of-home spatial activity, and life satisfaction. However, they also found that the correlation varied depending on gender and the type of spatial activity measures.

Although activity theory has been used widely as a theoretical reasoning in the literature, it has not been fully proven empirically (Isaacson et al., 2017). The theory does not state explicitly what types of activities would be beneficial, but rather that any activity is positive. Nevertheless, many activity types have been proven to be beneficial for health. Therefore, looking into the types of spatial activity might show a relation between spatiality and health.

Older adults' mobility is affected by many factors, such as gender, age, cognitive impairment, and car access (Habib and Hui, 2017; Shoval et al., 2010). Wong et al. (2018) also added multiple public transportation measures, such as seat availability, travel fare, and walking and waiting times. Marquet and Miralles-Guasch (2015) proved that the walkability of the neighborhood environment increases the duration and frequency of walking in the aging population. Mollenkopf et al. (2004) concluded that: "a person's physical, economic, social and technical resources as well as the structural resources prevailing in the area in which he or she lives in, are decisive preconditions of out-of-home mobility." The ability to mobility increases also the quality of life for the aging population compared to being housebound (Farquhar, 1995).

There is a clear link between mobility and health, as it is by definition' ability to move', and thus part of the overall health. However, examining the GPS trajectories of a person does not only measure their mobility but also their behavior, which is also associated with health. For example, algorithmically processed sensor data of social contacts, time spent outdoors, physical activity, and travel habits have been associated with behavior (Chaix, 2018).

Increased age is generally associated with reduced spatial activities (Bond et al., 2007). The mobility generally decreases with increasing age due to sensory and physical impairments, together with socio-physical barriers (Mollenkopf et al., 2004). However, this relationship is highly complex and varies between individuals, groups, environments, and other factors (Schwanen and Ziegler, 2011).

Different types of places and activities have been found to affect health. Places are locations with affordances attached to them, where affordance is the function that the place provides (Gibson, 1979). For example, green and blue places have been found to have multiple health benefits, such as lower psychological distress (Nutsford et al., 2016; van den Berg et al., 2019), stress, depression, and anxiety (Cox et al., 2017), and higher levels of concentration (Ottosson and Grahn, 2005). Personal preferences are likely to have an impact on the restorativeness of the environment. Park visitations have been found to be impacted by the personal nature orientation more than the proximity to a green space (Lin et al., 2014). The perceived restorativeness of the environment is impacted by personal preferences (Wilkie and Stavridou, 2013). The perceived biodiversity can have a higher impact on wellbeing than actual number or species (Dallimer et al., 2012). Therefore, it is plausible that the impact on health by different environments is impacted by personal views as well. Microscale elements of urban structure can influence the usability of the environments (Ottoni et al., 2016). Supportive social relationships are especially important for predicting happiness in the aging population (Baldassare et al., 1984). The subjective quality of social relationships has been one of the strongest positive correlators of well-being in the aging population and activity types (Litwin and Shiovitz-Ezra, 2006).

2.2 Place visitation from GPS trajectories

2.2.1 GPS trajectories and semantic enrichment

Movement is one major type of spatial activity. GPS trajectories represent the movements of entities in real world (Xiang et al., 2016). They allow for examining activity patterns in certain localities and complex behavioral models (Kwan and Lee, 2003). Trajectories can be enriched with much semantic information. One way is segmenting into binaries of 'stop' and 'move' series. Stop episodes consist of a consecutive series of a person’s GPS waypoints labeled as stop points in one location. The stop episodes are likely to be associated with meaningful activities (Parent et al., 2013). Another way is to use external geographical context such as points of interest (POIs) for enriching the trajectories and the outputs of the stop-move detection.

Stops represent the entity's stay in a certain location (Luo et al., 2017). A stop can be defined in multiple ways, which will affect the final allocation. Many studies include measures such as speed, time, location, heading change and density, for their definition and thus for their detection methods (Buchin et al., 2011; Hwang et al., 2018; Krueger et al., 2015; Luo et al., 2017). This research will be guided by detecting meaningful stops for health research.

A common way to enrich trajectories semantically is to use stops with an assumption that at a stop the subject is performing some activity in some place (Xiang et al., 2016). When the stops are then attached to the underlying geography, they become meaningful. Stops can be classified as active and non-active, e.g., waiting at traffic lights would be a non-active stop (Gong et al., 2015). However, non-activity can be quite subjective in some cases. Thus, it can be more useful to classify stops as important or unimportant, depending on the purpose of the analysis (Bermingham and Lee, 2019). It is also possible to use other information than stops, such as the travel mode, time in transportation, or the form of the trajectory (Krueger et al., 2015; Lehmann et al., 2019; Ying et al., 2010) to enrich a trajectory semantically.

2.2.2 Stop detection from GPS trajectories

The data used for the stop detection varies. With the current increase in mobile devices, GPS trajectory data is commonly available, and therefore many studies have been examining stop detection methods based on purely GPS data (Gong et al., 2015). This has its advantages as the availability of other sensor data, such as accelerations and recordings, vary depending on the sources.

Stop detection can be challenging due to multiple factors. The GPS devices have imprecision in the measurements, even in the best conditions (Trajcevski, 2011). The noise only worsens with a cloudy sky, nearby tall buildings, or other obstructions. Stop detection methods with specific parameters to detect stops, such as minimum time duration and maximum distance, e.g. (Biljecki et al., 2013; Schuessler and Axhausen, 2009; Stopher et al., 2008), are particularly sensitive to noise (Hwang et al., 2018). Gong et al.

(2015) developed a method that considers direction change, time, and density. However, density-based algorithms are prone to inaccuracies due to signal loss or noise in sparse trajectories. Hwang et al. (2018) argue that noise in the trajectories can be further combated by using linear interpolation and fuzzy interference through speed, angle, and circuity.

Another method that combats noise in a trajectory while detecting stops is a moving window (kernel) approach that determines the homogeneity within the kernel (Das and Winter, 2016; Wan and Lin, 2016). A moving window is a series of observations that are used to subset the trajectory iteratively. With this approach, it can be difficult to accurately determine the stops with an arbitrarily defined window. This can be further combated by detecting a suitable window size from the trajectory automatically (Bermingham and Lee, 2018).

Depending on the entities studied, the stop definition and optimal parameters vary. For example, a stop with a car will have higher acceleration and higher speed when moving, compared to walking. A person on a bus will have small stops along the way as the passengers get in and out, but the specific person can remain in the transportation mode. Thus, the stop detection method usually depends on the data used. This makes the analysis especially hard when the dataset contains different types of movement. Waga et al. (2012) aimed to detect different movement types and stops and highlighted the different speed profiles between the move and stop episodes based on the activity type. Therefore, many stop detection methods focus on one or few transportation modes to simplify the detection process, e.g., for buses (Garg et al., 2018), car, bus, truck, and taxi (Alvarez-Garcia et al., 2010), or taxis (Xiao et al., 2013).

Stop detection methods can be divided into two main categories: static and dynamic methods (Bermingham and Lee, 2018; Gong et al., 2015). In static methods, places are first derived, then trajectories are matched to these possible locations, e.g., Alvares et al. (2007). However, this is computationally very expensive or needs some previous knowledge of the probable stopping locations, which is most often not available. Dynamic methods, on the other hand, use the spatio-temporal attributes of the trajectories to categorize the trajectories to stop and movement episodes (Luo et al., 2017). The dynamic methods can be further divided into speed- (Palma et al., 2008), clustering- (Lv et al., 2016), centroid-, density- (Gu et al., 2017), or hybrid-based methods (Gong et al., 2015). However, it is uncommon to use these methods in complete isolation as none of them is solely sufficient to detect stops accurately.

A hybrid model between the static and dynamic methods can also be used. Often some information, such as opening times of the stores, can be known, which can be also linked to the probability of visitation. For example, Luo et al. (2017) used DBSCAN and incorporated automatic detection of a density threshold and movement ability. They relied on an assumption that during a stop episode the distance between the first point and the last point of a sub-trajectory segment, is considerably smaller than the sum of all distances between the points of the segment.

Commonly a dynamic approach uses the density-based clustering to detect clusters of waypoints, which is a likely stop, e.g. (Hwang et al., 2018, 2017; Luo et al., 2017). However, Bermingham and Lee (2019, 2018) argue that these methods require many user-defined parameters, such as maximum speed and minimum time. The selection of these parameters is not a simple task as optimal numbers might vary depending on the stop definition, transportation mode, etc. Therefore, they propose a probabilistic stop detection method (POSMIT), which factors in the trajectory's sequential nature.

A common problem of the GPS data is also the noisiness of the data. When there is noise, this can suddenly increase the speed, decrease the density, and increase the displacement in a short time, which will often lead to a false classification of a move. There have been many attempts to combat this uncertainty issue (Gong et al., 2015; Hwang et al., 2018; Tran et al., 2011; Xiang et al., 2016). POSMIT aims to combat this by using a moving window that calculates the probability of a waypoint being a stop based on the other points and the variance parameter.

2.2.3 Semantic enrichment on stops for modeling place visitations

The stop episodes' semantics can include different types of data, such as a person's feelings, duration, conducted activity, etc. (Krueger et al., 2015). For example, semantic data derived from GPS trajectories are used to detect wandering behavior (Lin et al., 2012), optimize traffic in a city (Brakatsoulas et al., 2004), detect personally semantic places (Lv et al., 2016), provide personalized place recommendations (Ruta et al., 2012, p. 20), and infer travel behavior (Shoval and Isaacson, 2007; Zheng et al., 2009). In this research, we will mainly focus on the place types extracted from Points of Interest (POI). This method of semantic enrichment can also be called place-matching (Bermingham and Lee, 2019). This is a common method often used when dealing with GPS data.

GPS data is commonly used to connect places to stop episodes. In comparison to a travel diary, GPS data is generally easier to collect. In travel diaries there can also be inaccuracies, such as reported arrival and departure times or entire stops missing (Bohte and Maat, 2009). There is an increasing use of GPS in semantic enrichment in health research as it has only a small burden to participants, allows for geographic temporal assessment, personal exposure to areas, and the algorithm-based identification of visited trips and places (Chaix, 2018).

Databases of POIs and their labels are often used to extract the semantic meaning. The sources for geographic data include multiple sources, such as Google Places API (Lv et al., 2016), Foursquare (Gu et al., 2017; Krueger et al., 2015), and AlpsMap (Gong et al., 2018a). Many papers use OpenStreetMap data (Bermingham and Lee, 2019; de Graaff et al., 2013; Ruta et al., 2012), because it includes user-generated semantics as place types and can thus be argued to include commonly accepted semantic meanings.

The place-matching methods from GPS trajectories based on the underlying geography can have many issues. For example, stop episodes associated with multi-story buildings or places close to one another make the place detection often unreliable as the stop episodes are composed of probable stop points (Chaix, 2018). Hence, multiple methods for place-matching have been developed.

Many semantic place-matching methods to form stop episodes are based on concepts of machine learning, such as Hidden Markov Model (HMM) (Bermingham and Lee, 2019; Lv et al., 2016; Shi et al., 2019a, 2019b, p. 1549-1557; Yan et al., 2011, 2010) or random forest (Gong et al., 2018a), hierarchical decision tree (Fileto et al., 2013; Krueger et al., 2015), proximity to geometry (Alvares et al., 2007; de

Graaff et al., 2013) or probability (Celli et al., 2010; Gu et al., 2017). Many models use a combination of methods, e.g., Gong et al. (2018b) developed a model with DBSCAN-TE and support vector machine.

Deep learning methods are used increasingly in POI detection, but they have inherent issues depending on the post-analysis. Feng and Timmermans (2016) compared naïve Bayesian, Bayesian network, logistic regression, decision table, SVM, multilayer perceptron, and C4.5 algorithms and reached good results with the Bayesian network. HMM models can suffer from data scarcity, which can be combated with state-sharing of some of the data with all the users while maintaining some of the latent personal states (Shi et al., 2019b, p.1549-1557). A dynamic Bayesian network includes the temporal aspects of the model and has been used increasingly in the semantic enrichment of places (Meng et al., 2017) for ship trajectories: (Wen et al., 2019). Meng et al. (2019) tested different place inference models and concluded that out of K-nearest neighbor, support vector machine, artificial neural network, and random forest and dynamic Bayesian network (DBN), and found that the DBN outperformed the other models in most of the place types based on F measure and accuracy. They also tested if social media popularity measures would improve the detection accuracy and found that it improved the detection of certain places but decreased accuracy of other places such as ‘Recreation’ and ‘Transportation’. Probably because these places do not get as much attention on social media.

The deep learning methods usually incorporate personal information, the trajectory, or overall visitations to the POIs. However, if the post-analysis includes comparing different POI visitation patterns between people and trajectories, this could create a bias. These methods could shift the visitations to more mainstream in terms of all POI counts and leave out the less-visited locations. When considering the other visited POIs by the same person, it could create a bias for the person to visit more similar POIs in the model than in reality. One study by Shi et al. (2019a) divided people into four groups based on the diversity of visited POIs and found that the HMM performed worse with people with high diversity in POIs. In most of the studies, the goodness of the POI association model is measured through F-measure, recall, precision, and accuracy. However, these results do not show whether the results are biased and to which magnitude.

In the case of post-analysis with these measures included, more physical approaches can be used. However, it has been found that simply taking the nearest POI can lead to average results. However, these results can greatly improve through many methods (Krueger et al., 2015). Krueger et al. tested different methods and found that factoring the visitation counts and distances improve the results significantly.

Geometry based methods embed the places to the trajectories based on whether they intersect spatially. For example, Alvares et al. (2007) developed a SMoT to detect places based on the spatial patterns on the overlaid geometry, which was further expanded by Moreno et al. (2014) by taking into account the hierarchy of nested geographies. To combat the missing GPS values inside buildings, the GPS points right before and after the stop episodes have been used (de Graaff et al., 2016).

A hierarchical decision tree can also be used to match the places. Krueger et al. (2015) used the counts of the POI types to determine the most likely visited place type. Their model takes the assumption that if there are many similar places nearby, it is more likely that the person visited one of those places. They also note that it could be beneficial to take into account the average distances to the different places.

2.3 Place modeling

Place in geography is the subjective meanings added to a space, whereas a location is the exact point with longitude and latitude. Gibson (1979) states that the different relationships to a place are due to affordances. The theory of affordances argues that objects, such as places, are seen through what they provide for humans. For example, a library can provide a place to read and a school provides a place to learn. It has been shown that locations are stored cognitively by functional groupings of all kinds (Mark et al., 1999). Just as locations are connected to each other by groupings also their effects on health and wellbeing can be defined through different groupings.

Another view is the humanistic view that incorporates the understanding of a subjective place into the research (Hummon, 1992; Tuan, 1979, 1975). Humans attach places with a sense of place which is constructed through human behavior, the physical environment, and social or psychological processes. It includes attachment, meaning, and satisfaction towards a place (Kyle and Chick, 2007; Stedman, 2003). This is influenced not only by the personal relationship with the place, but also through collective identity construction (Greider and Garkovich, 1994; Stokowski, 2002). The affordance theory can be seen as a practical part of the sense of place.

A place can be determined in multiple levels of granularity, which determines how specific a place is. Detail and extent determine the level of granularity with higher granularity being more specific and smaller scale (Vulchanova and Zee, 2013). Granularities can be defined as classes of spatial and semantic. Spatial granularity handles the metric granularity, whereas semantic granularity uses the meanings of the entities that represent the real world (Fonseca et al., 2002).

Different granularity levels in trajectory modeling can influence the results. The places that an individual has occupied in a certain time frame can vary depending on the granularity. In coarser granularity, an entity might be staying in one place, whereas in a finer granularity level, this might be divided into multiple smaller places (Hornsby and Egenhofer, 2002).

2.4 Place clustering

Places can be clustered from a fine granularity to a coarser granularity based on their spatial, temporal, or semantical similarities. Spatial clustering can be used to identify places that are linked to other places. According to Tobler's first law of geography, "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). Hence, places that are closer to each other are also similar to each other. Spatial clustering uses most commonly Euclidean distances to measure the differences between locations. Clustering methods can be divided into partitional, hierarchical, grid-based, model-based, and density-based methods (Birant and Kut, 2007). Methods for spatial clustering include the partitioning method such as K-Means, density-based such as DBSCAN, and hierarchical algorithms such as BIRCH.

In addition to spatial clustering, spatial-semantic clustering has been explored in research. The methods for the semantic clustering are the same or similar as for the spatial clustering. Deng et al. (2009)

used DBSCAN to cluster tags with spatial reference to receive semantic information about the location. Shi et al. (2014) extend DBSCAN for social networks also by considering both the spatial and social distances and found that it could find social clusters more accurately rather than purely spatial clusters. Steiger et al. (2016) investigated it using multi-dimensional neural-network-based 'self-organizing maps' to analyze semantics of geotagged Twitter data and found that there were semantic and spatiotemporal similarities that could reveal latent urban structures.

DBSCAN works well when the data is low dimensional and the quantifying parameters are known, so that the density threshold can be effectively defined. This, however, is always not the case. High dimensional data sets often have a skewed distribution of the data and thus do not scale well for global parameters. For semantic clustering, the vector distances have to be derived from another parameter, rather than Euclidean distance as this loses its meaning in high dimensional data. To combat this challenge, OPTICS was developed by Ankerst et al. (1999) to further improve DBSCAN. Inglese et al. (2017) used OPTICS successfully for tumor detection from high dimensional data. Nimmo et al. (2017) found that although PCA combined with DBSCAN is most used in partial discharge activity in power networks, t-SNE with OPTICS is more successful at clustering the data. These examples show that although the combination of t-SNE and OPTICS has not been much used in research, the results are positive. However, as T-SNE and PCA reduce the dimensionality, they will also lose some of the information.

A way to combat the issues of high dimensionality and the loss of meaning in Euclidean distance is through hierarchical clustering with cosine similarities. In hierarchical clustering, each data point forms in the beginning its own cluster and the clusters are then combined iteratively with the method chosen for the distance measure. This does not require dimension reduction, but as each data point forms initially its own cluster, it does not scale well to large data sets but works well on relatively small data (Pandove et al., 2018).

2.5 Embedding and place modeling

Embedding machine learning can help to understand places spatially and semantically. Embedding in machine learning is a technology that enables modeling of discrete categorical variables as continuous vectors. Because places are a combination of both spatiality and semantics, it is often necessary to study them together. As computers do not recognize semantics naturally from human language, vector space models have been applied for decades to enrich the word information with meaning (Turney and Pantel, 2010). These models are based on the idea that words that occur in similar contexts are also similar to each other (Harris, 1954), which is a very similar idea to the first law of geography.

Place2vec was developed by and Yan et al. (2017) and is derived from Word2Vec developed by Mikolov et al. (2013). Place2vec uses the visitation counts and geographic aspects to embed the semantics. It transforms the place semantics and the actual distances into a high dimensional vector space to combine the holistic similarities between places. The distances between the places can then be used to analyze the similarities between places. This model, however, requires a lot of visitation data so that each POI has multiple visitations, and if the study area is large, this can also become computationally expensive.

There is emerging research concerning behavioral models based on places. One example is to detect unusual patterns in normal visitation areas (Tung et al., 2011). Ying et al. (2010) used the longest common sequence between annotated semantic trajectories to find similar people. Ying et al. (2012) used decision tree and SVMs to build a multi-level classification model to predict the demographics of the user based on their mobile usage. This model was then further developed to predict the next semantic place (Huang et al., 2012). Cao et al. (2020) developed habit2vec to find similar people based on their habits. Habit2vec embeds the POI visitations for each person with other POIs during a similar time period to derive context. They then derive a habit vector for each user to then cluster people based on similar habits. A semantic-aware HMM was developed to group people based on similar mobility patterns by Shi et al. (2019). They applied POI type, spatial grid, and temporal hour windows, used this information as latent states for HMM and von Mises-Fisher to cluster similar people based on the cosine similarity.

There are an increasing number of ‘trajectory to vector’ models developed successfully to model places and their similarities. They generally aim to model places based on the sequential nature of the trajectory. Zhang et al. (2020) developed Traj2Vec and classified urban areas effectively based on mobile trajectory data. Bin et al. (2020) used Traj2Vec to develop a model to simulate user behavior and give recommendations of places and found their model to outperform the state-of-the-art. Mot2vec was developed by Crivellari and Beinat (2019) primarily to group similar places together based on trajectory behavior but can also be used to compare similarities between users. Crivellari and Beinat examined the performance of Mot2vec on tourism data and found that the model can provide meaningful semantics of the places and that the visitation patterns can be associated with the tourists’ country of origin when the averaged vectors were visualized with t-SNE.

2.6 Research gap

In this study, we use semantically enriched trajectories to investigate the relationship between POI visitation patterns and health indicators. There is increasing research on different POI patterns and their modeling, and there have been studies with spatial activity comparisons to health, but to our best knowledge, this is the first study to attempt to combine the two by examining health through POI visitation patterns with GPS trajectories.

3 Research objectives

Places have different affordances depending on their semantics and location. Therefore, they may also have different implications on health. By studying participants’ trajectories, we may model the affordances of the places. According to the activity theory, the total number of visited places would be associated positively with overall health, although this has been challenged in literature. Therefore, we anticipate that the total place visitation could be an indicator of mobility and consequently be associated with physical health. Based on previous literature, we expect that some clustered place types, such as green spaces, are associated positively with physical and mental health. In addition, places that have a social setting could be associated with higher mental health. Physical health is likely to be associated with clusters that indicate exercise. Through the place clustering, we can see whether certain place types can have a connection with health in the older population. The data is analyzed through five steps (Figure 9).

The GPS trajectories are segmented into stop episodes and enriched with semantics from POI types. The stops are then clustered based on their semantics and the visitation order. The multivariate regression analysis shows the relationship between older adults' health and the visited places. These steps will be further elaborated in the following sections.

4 Data

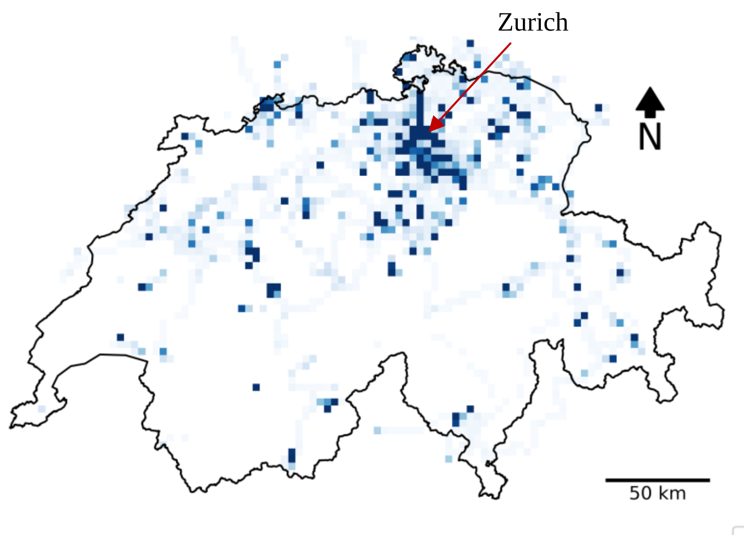

Figure 1 The distribution of the observations. The darker shade represents more observations with a maximum value of 1 million within a cell to represent density. Observations within the bounding box of Switzerland are included in the image.

The data consists of trajectories of 159 people of over 65 years old. Only participants with reasonable mobility were included in the study. Each of the participants carried a GPS logger (i.e., a uTrail device) for about 30 days. These form the GPS trajectories, which include longitude, latitude, timestamp, speed, number of satellites, altitude, etc., with a sampling rate of 1Hz. Occasionally the records can have longer time gaps in case of signal loss. The data was stored in the SD card in the uTrail device and downloaded periodically in two weeks intervals. Participants carried the uTrail device attached on a pocket or a waistband and charged them overnight. The participants gave written consent to the usage of their data. Most of the GPS observations are from the Zurich region as the study was conducted in Zurich and required participants to travel to an in-person meeting (Figure 1).

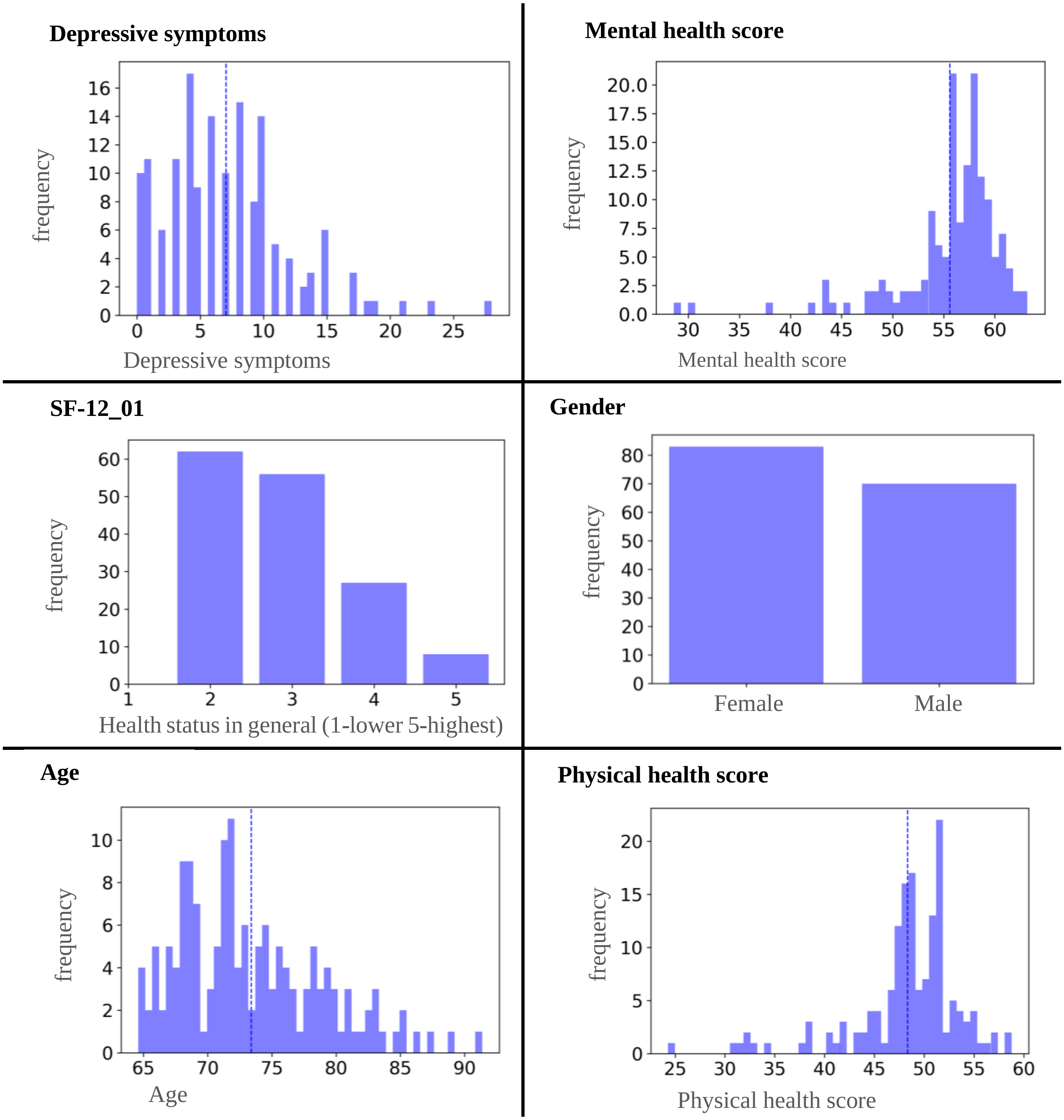

Figure 2 Distributions of the health indicators

Participants were surveyed for physical and mental health during the study using different questionnaires, such as short-form health survey (SF-12) and depressive symptoms score (ADS). The health indicators used in this study are the Depressive Symptoms Score, Mental Health Score, Physical Health Score, and Self-Assessed Health Score (SF-12_01). We also included age and gender in this study as control variables. The histograms of the survey results can be seen in Figure 2. The nature of the project required the participants to have a minimum level of cognitive and physical abilities, which would have affected the distribution of the data. This can also be observed as a slight positive skewness in the health indicators and as negative skewness with age (Figure 2).

The distribution of observations over time seems to be reliant on the number of participants during that day (Figure 3 and Figure 4). The distribution over the weekdays seems to be highest at the beginning of the week and decrease towards the weekend (Figure 3). There does not seem to be a particular season that would have a higher amount of observations, but rather the observation fluctuations are due to the number of participants (Figure 4). The period of study is between April and November.

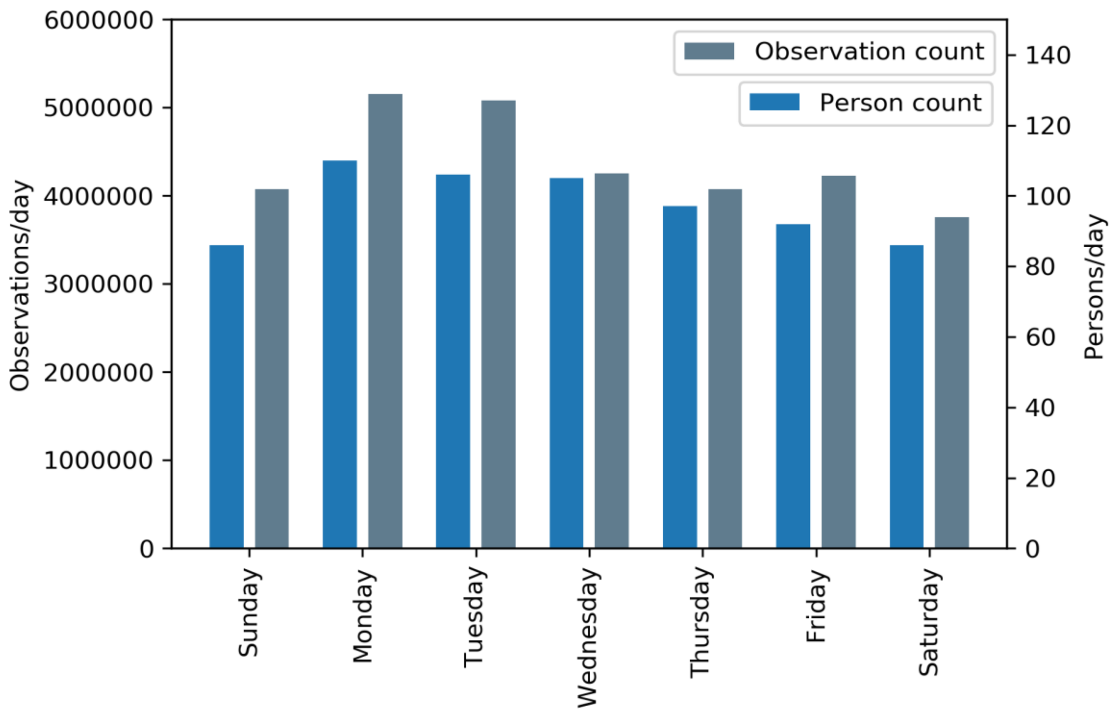

Figure 3 The distribution of all the observations over the weekdays

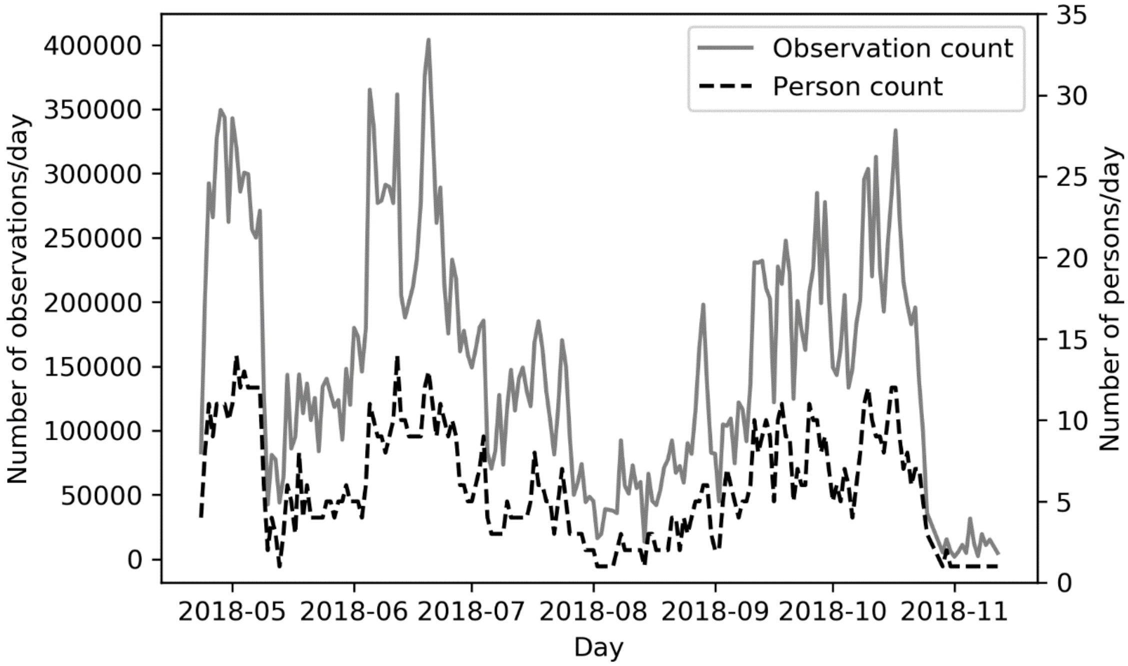

Figure 4 The observation count and the person count over the study period

OpenStreetMap (OSM) POIs were used as the resource of places to enrich the staying points. They include polygons and point features within the boundaries of Switzerland, derived from Geofabrik ("Geofabrik Download Server," 2020). OSM also includes other features, such as 'ways,' but as these are generally attached to movements rather than stops, we did not include them into the data. OSM data includes <"key"=>"value"> pairs for each geographical feature that we used for the analysis, e.g. <"landuse"=>"residential">. In a data frame structure, the 'key' can be either a column of a data frame or in the column 'other_tags' it can also be part of the cell. As the OSM data is user-generated geographical information, users can keep adding new customized key-value pairs continually to improve the tagging system. However, there are informal standards such as "Map Features – OpenStreetMap Wiki" which we used as a guideline for the most common tags (Appendix section A).

4.1 Data pre-processing

GPS data usually includes some noise that needs to be cleaned for specific purposes. We removed the days that do not have GPS data for a minimum time span of six hours and 1000 observations. The observation count ensures that there is a minimal number of gaps within the 6h duration. This resulted in the number of participants to 124 and the valid days to 1193 out of 1506 with any GPS data. The points outside Switzerland were removed to ensure the POI structure to be similar. The qualified days were then rechecked if they still qualified after the removal. The number of days was thus further reduced to 1186 in total with a time span of 23/04/2018 – 12/11/2018. Most of the fluctuation in the observation count is accounted for by the number of people recording (Figure 4). No participants were removed in the second round. Later in this thesis, the whole dataset refers to this cleaned version. As we were mainly interested in stop episodes, data points with speed of over 70m/s were also removed as noise as this is the highest train speed in Switzerland.

For the statistical analysis, we sampled a subset of the data where all the participants have the same number of days to ensure that the POI frequency is not increased in certain participants due to higher availability in GPS data. We compared the number of days/number of the participants in Figure 5. We tested minimum days of 6 and 12, where the six-day threshold resulted in participants of 88, and the 12-day threshold resulted in 53. The statistical analysis was therefore conducted on the six-day random sample to maintain variety in the participant sample.

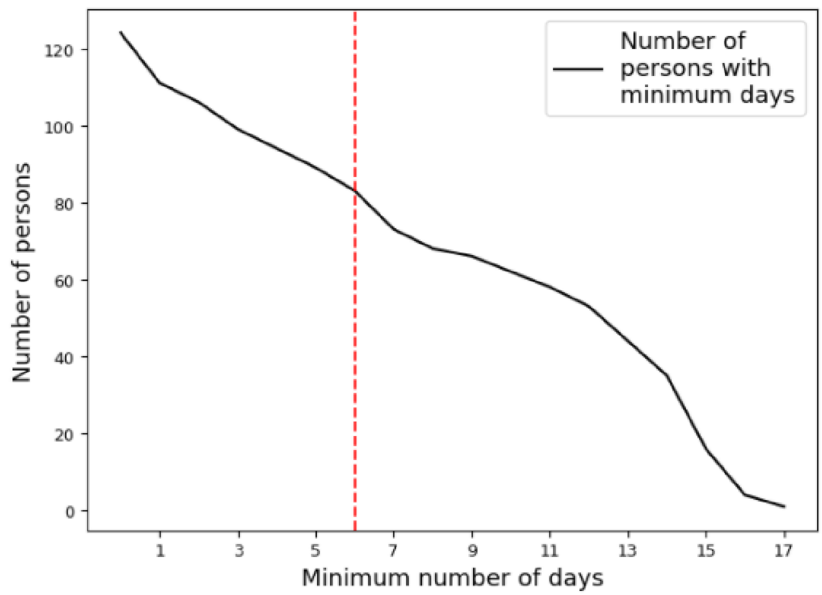

Figure 5 The number of people with minimum number of qualified days

The OSM data includes data that is not needed for this analysis and was filtered out. We identified the keys and tags that are important or unimportant for the analysis and filtered the data based on these. For the commonly used tags, we used the OSM wiki. These tags can be found in Appendix section A. As a base rule, we used the likely duration of visitation to filter out the irrelevant tags. For example, a trip to a 'bin' is likely to be very short in duration and is not included in our data as part of the potential POIs. We will discuss further in detail the important stop episodes in the section 5.3. After the semantic enrichment of stop episodes, we re-filtered the POIs as there were some POIs that were not in the wiki as common tags but appeared in our data.

Apart from the points and polygons, we also filtered a third class from the OSM data for the land use polygons. These are generally larger polygons with lower granularity. The filtered OSM data is illustrated in Figure 6, Figure 7, and Figure 8.

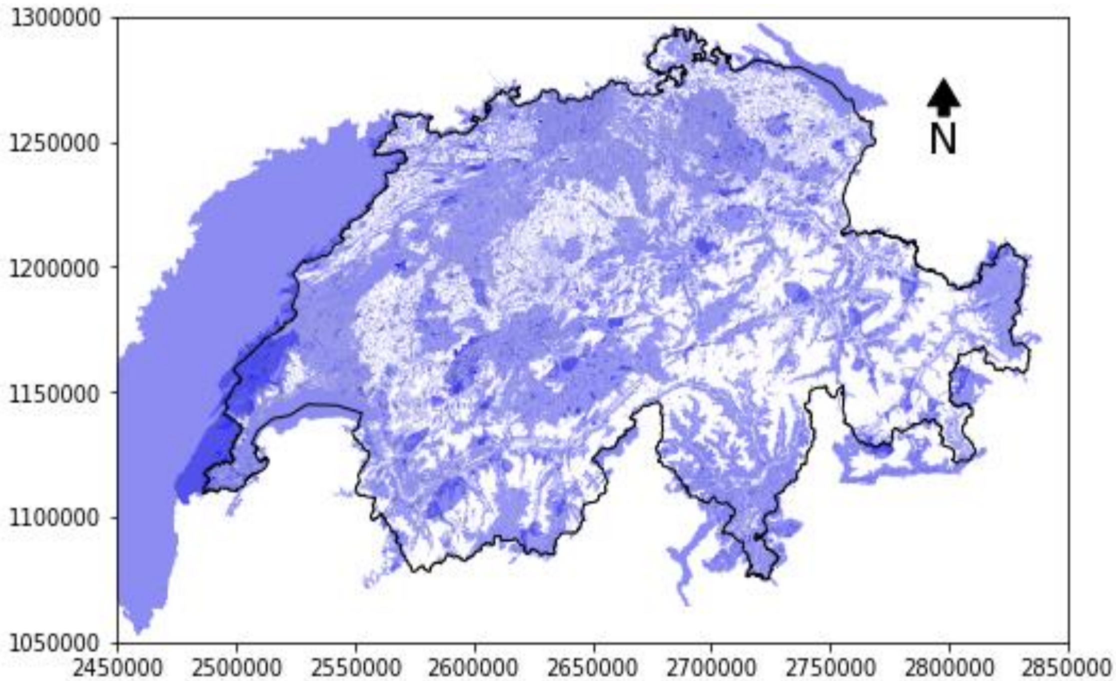

Figure 6 Polygon POIs derived from GeoFabrik for Switzerland and filtered for land use and nature.

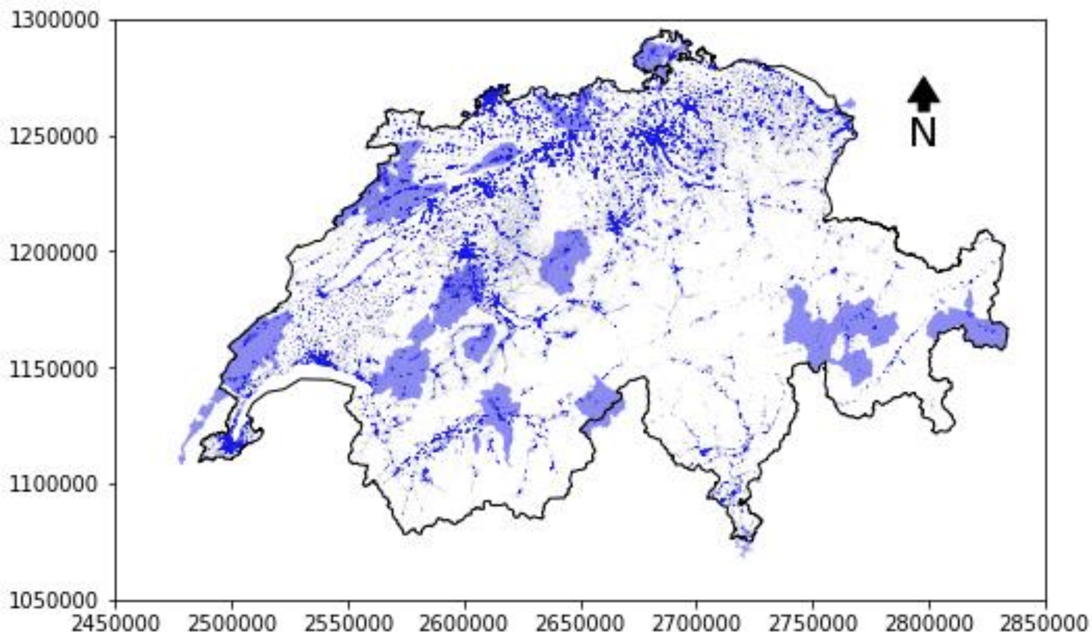

Figure 7 Polygon POIs derived from GeoFabrik for Switzerland and filtered

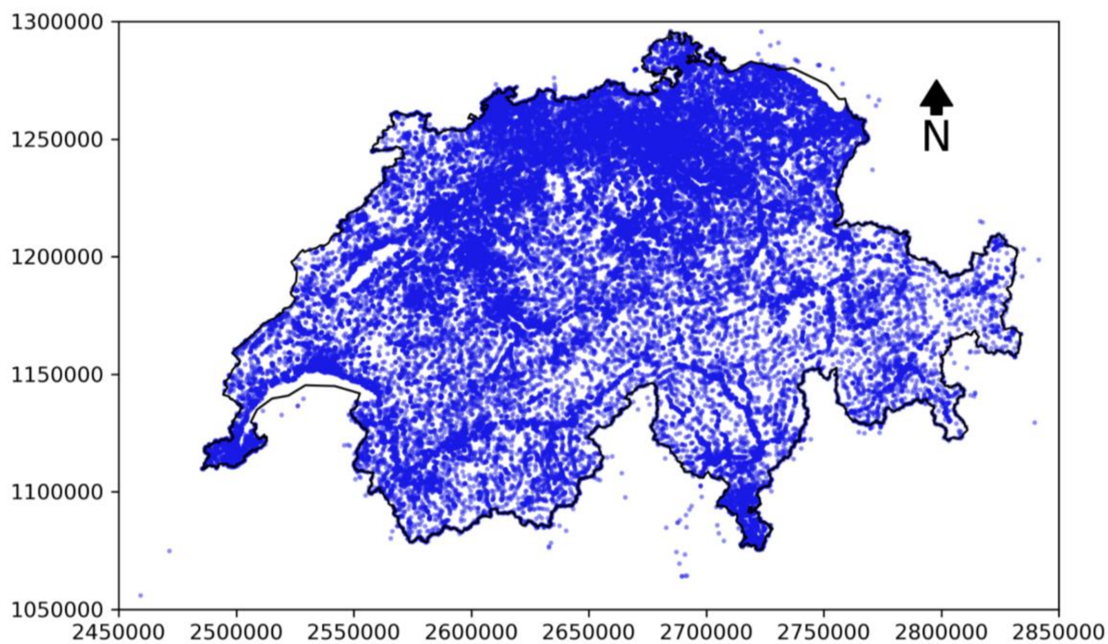

Figure 8 Point POIs derived from GeoFabrik for Switzerland and filtered

5 Methodology

5.1 Overview of workflow

First, a stop detection method was applied (Figure 9a) followed by the stop episode semantic enrichment (Figure 9c) with the OSM filtered POIs (Figure 9b). Place modeling enabled clustering that aggregated the places into semantically meaningful groups (Figure 9d). These groups were then used to test whether there is a relationship between health and place visitation (Figure 9e). The workflow of the methodology is illustrated in Figure 9 with the simplified workflow on the left side and the different methods expanded on the right side.

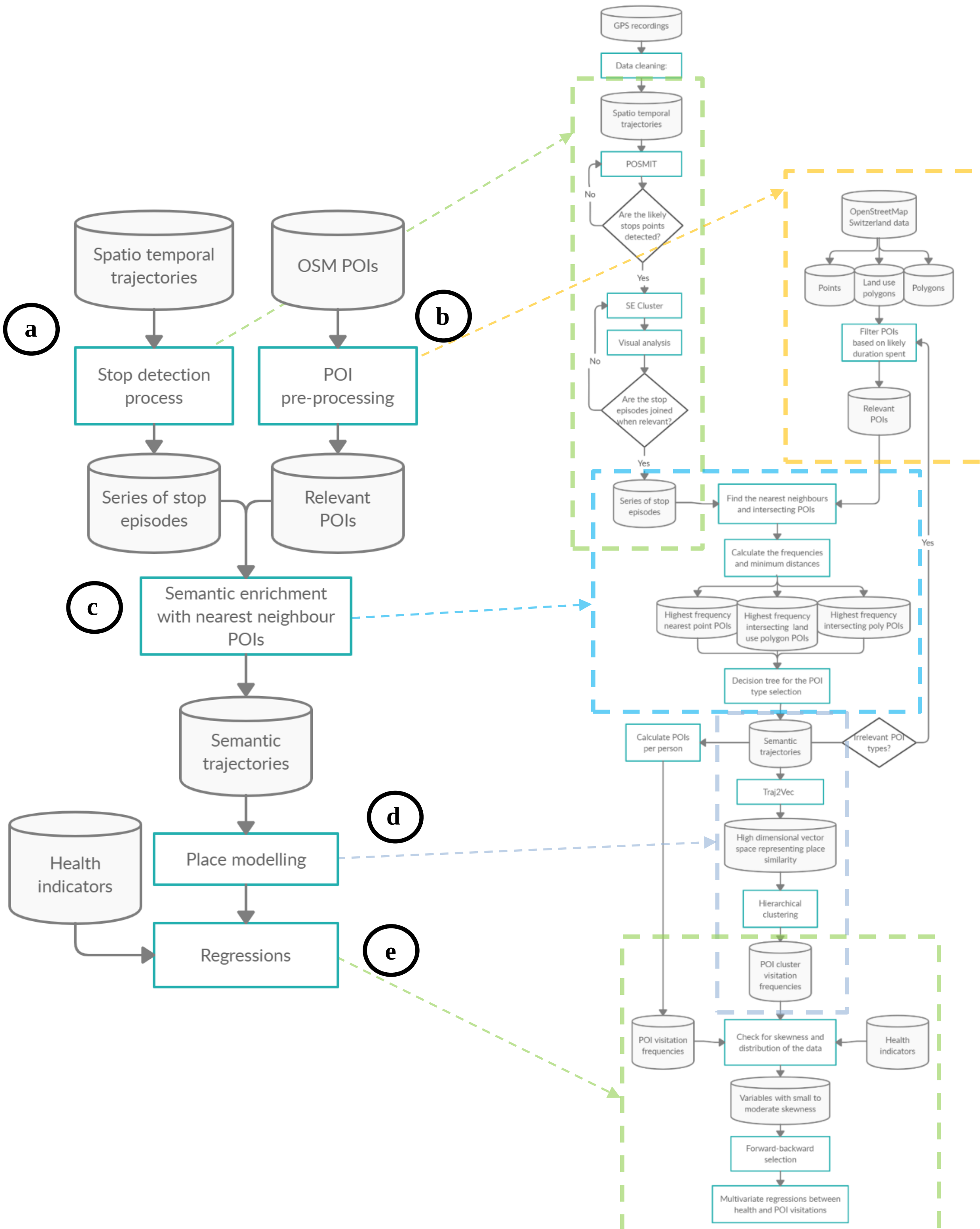

Figure 9 Workflow illustrated with different sections

5.2 Definitions

In this section, we define the concepts used in this thesis. We adopted our stop related definitions from Bermingham and Lee (2019, 2018), as we adapted their methods for the stop episode detection.

Definition 1:

GPS Observation Point, *P*: an observation in a trajectory with $< x_i, y_i, t_i.>$, where $x_i, y_i \in \mathbb{R}^2$, $t_i \in \mathbb{R}^+$

Definition 2:

Spatio-temporal trajectory, T_{st}: a series of P_i, where $i \in \mathbb{Z}^+$ and $t_1 < t_2 < \ldots < t_n$

Definition 3:

Stop point, P_{stop}: P_i, where $P_i \in E_{stop}$.

Definition 4:

Semantic point, P_{sem}: P_{stop}, with an annotated label a_i, i.e. $< x_i, y_i, t_i, a_i,>$, where $a_i \in A$ and $\forall\ P_{sem} \in E_{sem}$. A is the finite set of semantic labels based on the data mining results.

Definition 5:

Stop episode, E_{stop}: a homogenous segment in a trajectory that is entirely composed of P_{stop} that are spatially and temporally closely recorded, where $E_{stop} \in T_{st}$. E_{stop} is a set of chronologically consecutive P $<P_i>$ (start_idx < i < end_idx)

Definition 6:

Semantic episode, E_{sem}: E_{stop} that has been annotated with a semantic label, where $\forall\ E_{stop} \in T_{sem}$

Definition 7:

Semantic trajectory, T_{sem}: a series of semantic episodes.

Definition 8:

POI and POLOI: point of interest and polygon of interest, respectively, both will be referred to in this paper as POI for simplification.

Definition 9:

POI type, POI_{type}: semantic label of E_{sem} of POI_i

Definition 9:

Total POI count: the count of all the *POI*s, i.e. $|POI_{total}|$

Definition 10:

POI cluster POI_c: a set of POI_{type} that are semantically similar

Definition 11:

POI cluster count: the count of all the *POIs* in a cluster, i.e., $|POI_c|$

Definition 12:

Unique POIs: the count of the unique *POIs* within each POI_c, where *POIi* = semantic label of a E_{sem}.

5.3 Segmentation of the data into movements and stops

As stops are related to activities and these activities might be related to health, we focused on stop episodes and their underlying patterns. Our first goal was to separate the movements of the data to find the stop points (Figure 9a). Stop episodes are sub-sequencies of a trajectory, when there is little movement for a minimum amount of time. The movement periods were then discarded.

Identification of stop episodes from a trajectory is most commonly achieved through speed, centroid, time duration, density, or a hybrid method (Gong et al., 2015). In this research, we applied a hybrid method to detect stop episodes. First, we used POSMIT function (Bermingham and Lee, 2018) to detect the most probable stopping points by using spatial displacement as an indicator of the movement. It uses an index search bandwidth parameter (h_i) with a spatial stop variance parameter (h_d) to calculate the stop probabilities through a Gaussian kernel smoothing function which lowers the impact of noise. An index-based parameter is used instead of a temporal bandwidth to account for losses in GPS signals.

POSMIT requires three parameters: h_d, h_i, and minimum probability. We estimated the h_d by visualizing 20 sample trajectories' speed distribution and located the common elbow point (Figure 11). Bermingham and Lee (2018) included a method to automatically detect the h_i and the minimum probability from a trajectory, which was also applied in this paper. It is possible to estimate these based on trial and error. However, as our trajectories are very large and variable, we estimated them for each person, which allowed detecting better user-specific movement rather than basing the parameters on a few users or the average. The method for h_i calculates the sizes of the subsequences in the trajectory where the h_d is smaller than the defined parameter, takes the median of all the sizes, and divides it in half (as the h_i is used for both directions). The minimum probability is then calculated by clustering the stop probabilities calculated in POSMIT, to form two clusters based on K-Means. Then an average is calculated between the maximum value of the stop cluster and the minimum value from the move cluster. A sample trajectory illustrating the derived stop episodes through POSMIT is illustrated in Figure 12.

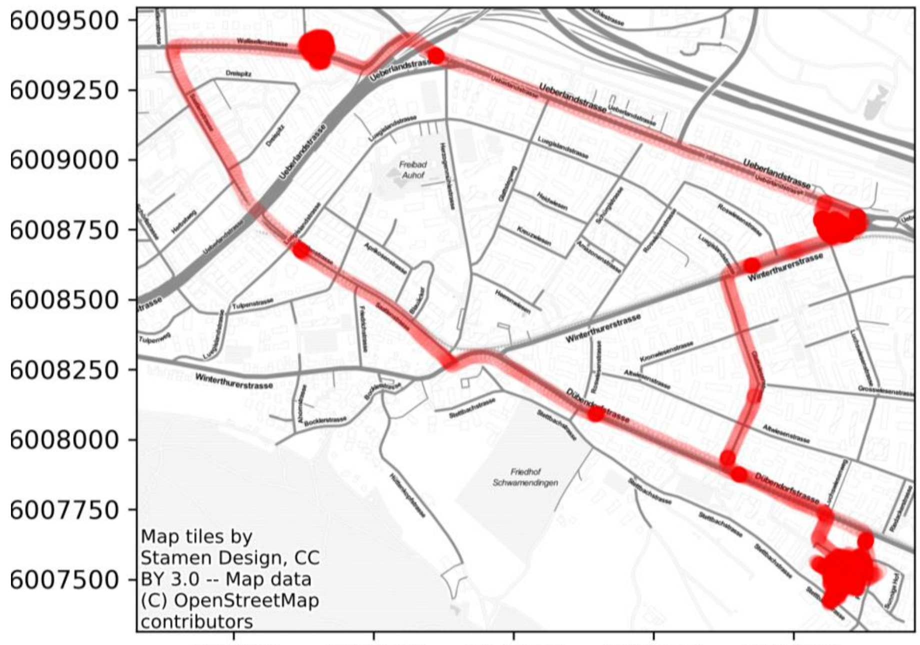

Figure 10 A sample trajectory. It shows three longer and wider stop episodes as well as some smaller denser areas that are most likely traffic lights.

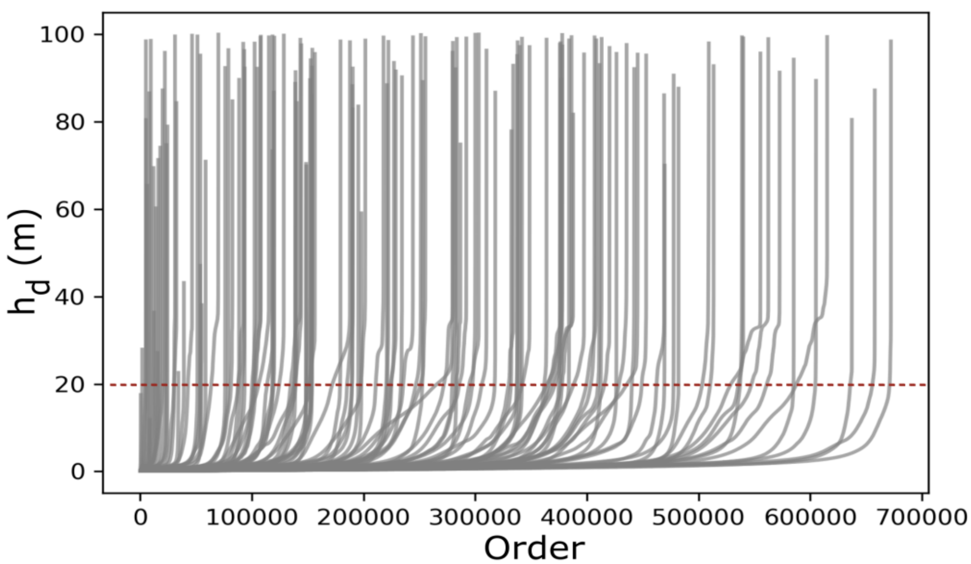

Figure 11 The distance distribution of the GPS points for each person. The elbow point is used to detect the h_d -value used in the POSMIT method. A common elbow point is detected and shown in the figure by the red dashed line at 20m level.

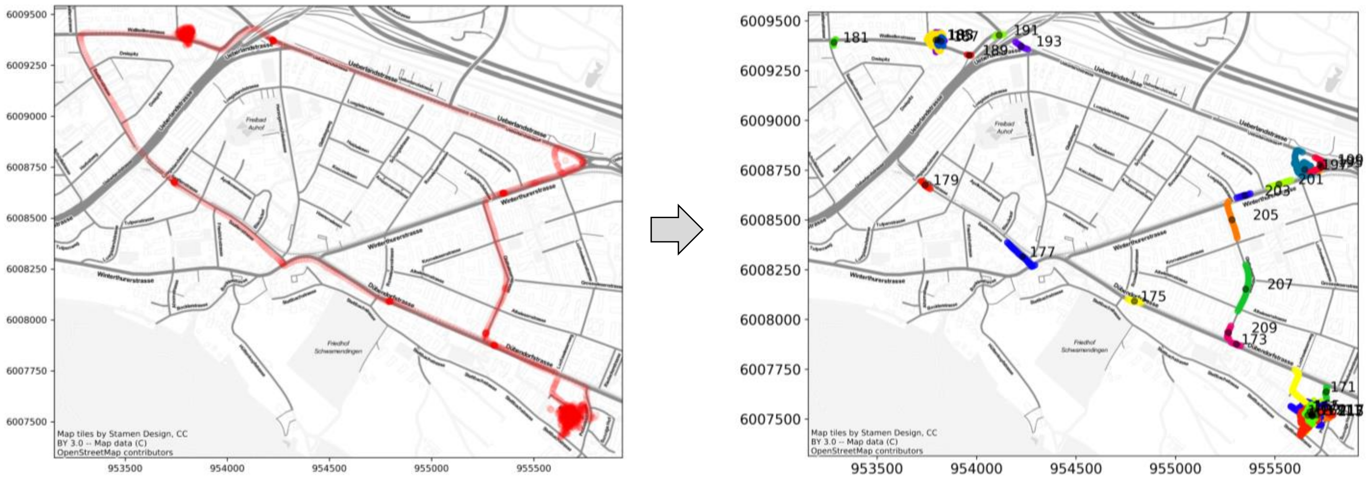

Figure 12 Transformation from trajectory to stop episodes with POSMIT

After POSMIT the stop episodes were further cleaned to derive the meaningful stop episodes for this analysis by specifying the minimum duration of 3 min and maximum speed of 3m/s. As can be seen in Figure 13, this cleans most of the ‘tails’ of the stops that occur at the beginning and at the end of the movements.

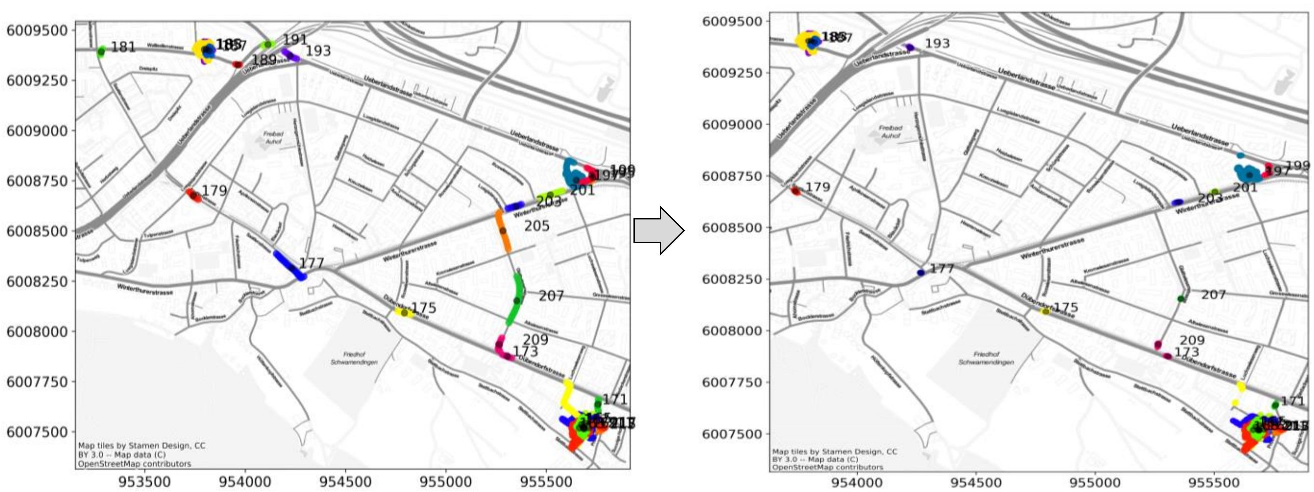

Figure 13 Stop episode cleaning with maximum speed

To successfully apply semantic enrichment, it is important that the stop episodes are sequences of t_{start} and t_{end}, where $t_{n-1\ end} > t_{n-1\ start}$ and without classifying one stop episode or multiple stop episodes. Especially with noisy data, the sequences easily get split as the algorithm detects outlier GPS points as movements. To combat this, we used an adapted method from Bermingham and Lee's (2019) “SeCluster” -method. We combined the stop episodes that are consequential and within 100 meters distance of another. As can be seen from a sample Figure 14 this combines many of the stop episodes that are spatially very close or even on top of each other. This is important in order to derive the correct frequencies for the later analysis.

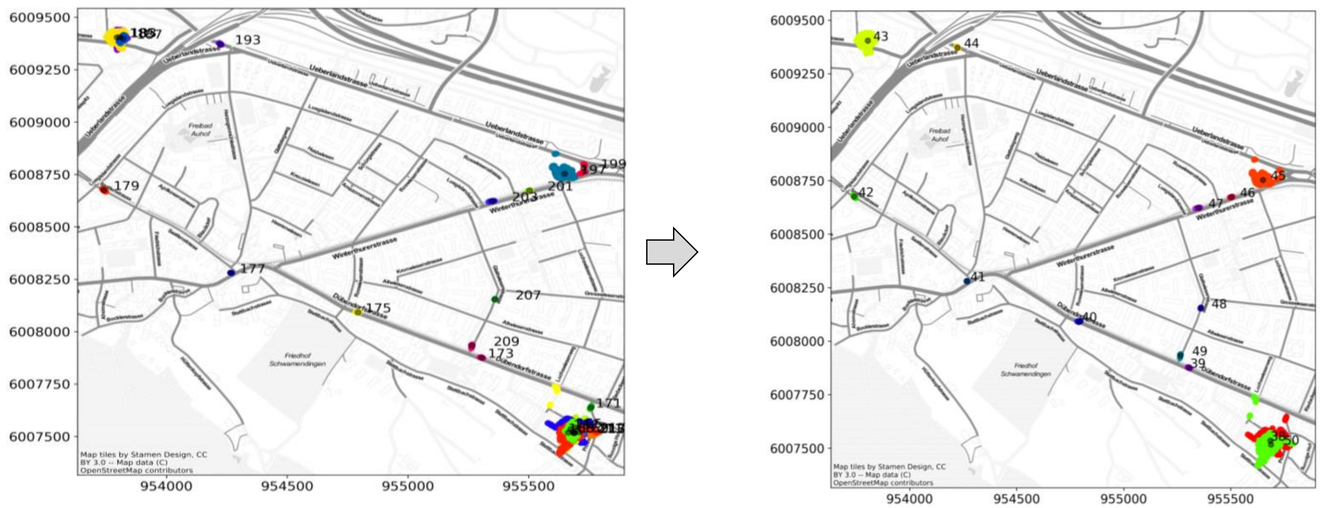

Figure 14 SE-clustering to combine stop episodes together based on consecutiveness and distance. Each individual stop episode is marked with a numerical label. From the bottom right corner and top left corner can be seen that there are multiple stop episodes directly on top of each other.

The minimum time periods for stop episodes vary between few minutes to hours (Parent et al., 2013). We used a 30-min minimum threshold that was also used in the approach by Bermingham and Lee (2019), together with POSMIT to derive meaningful stops. This excluded traffic-related stops that are not meaningful for the analysis (Figure 15).

Figure 15 Cleaning of the stop episodes based on minimum time. In this sample image we can see that most of the traffic related small stops are filtered out, as they are not considered meaningful,.

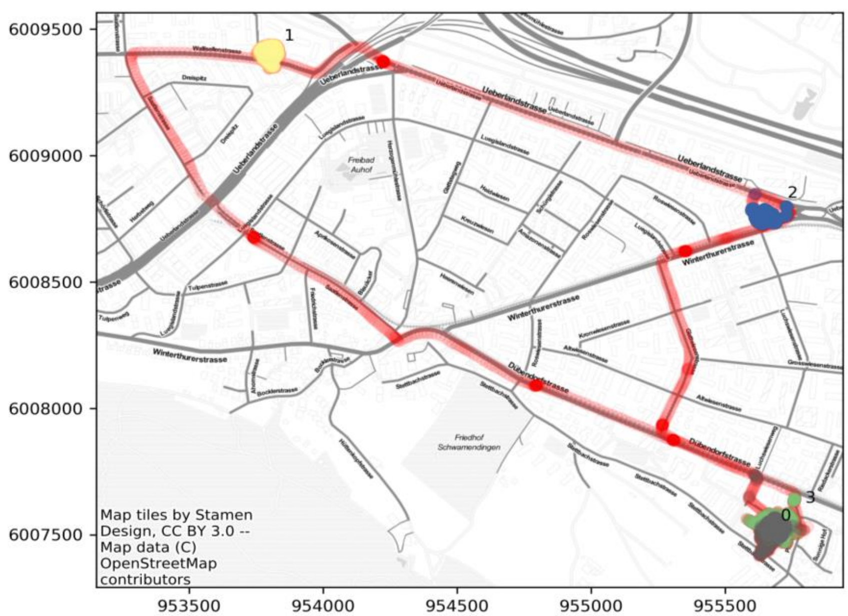

Figure 16 The final stop episodes for the sample trajectory with the consecutive labelling

Our parameter selection is motivated by selecting meaningful places. Hence, we aimed to initially detect even smaller stops and then further discard them based on whether they are unimportant for the analysis. We will discuss further in the following sections the selection of meaningful stop episodes as we discuss the enrichment of the stop episodes with the semantics.

5.4 Enriching the semantics of the stop episodes by POI types

Stop episodes are commonly used for trajectory semantic enrichment with the assumption that the stop episode includes an activity. The semantic enrichment of the activities from stop episodes is often done through the place visitation, e.g., if a person visited ‘restaurant,’ this can indicate the affordance of the place. However, the activities nor the semantics do not always provide useful information. For example, a person waiting at traffic lights is most often not important for the analysis but can be defined as a stop (Gong et al., 2015). As we discussed during the literature review, places occur in different granularities and entail an infinite amount of possible semantics. Therefore, we needed to extract the meaningful places with the most useful semantical information for our analysis.

During the pre-processing of the data, we separated the OSM POIs into points, polygons, and land use polygons and filtered them based on tags. As some places are not meaningful for most ordinary aged people, we needed to filter these out. We set the important semantic information based on their normal visitation period. A stop at a bin or traffic lights would be very short and not relevant to our analysis as this study focuses on the more prevalent place visitations. A bench, on the other hand, would have too small granularity for our purposes.

We used the filtered OSM data to enrich the semantics of the stop episodes. Figure 18 illustrates a small area with the point POIs and polygon POIs are shown with yellow dots and purple polygons, respectively. As can be seen from the image, there can be multiple point POIs within a polygon POI, and sometimes this could lead to problems if only point POIs were taken. For example, Figure 17 illustrates a possible scenario where most of the stop points are inside a building block. However, if only point POIs were taken, this would result in the store, rather than the building block. This could be acceptable if the trajectories entailed mainly these types of features, but since out trajectories include everyday life, with the participants’ residences, these types of features are important for the analysis. Polygons also include features such as ‘park’ and ‘farm’ which cannot be represented with a point. Figure 19 further illustrates some problems with taking the nearest neighbor of the centroid of the stop episode, as we would not associate these stop episodes with the joined POIs. Using a centroid can also lead to some problems, as centroids can be overly affected by ‘tails’ of the stop episode that occur at the beginning or at the end of the stop episode, as discussed in Section 5.3.

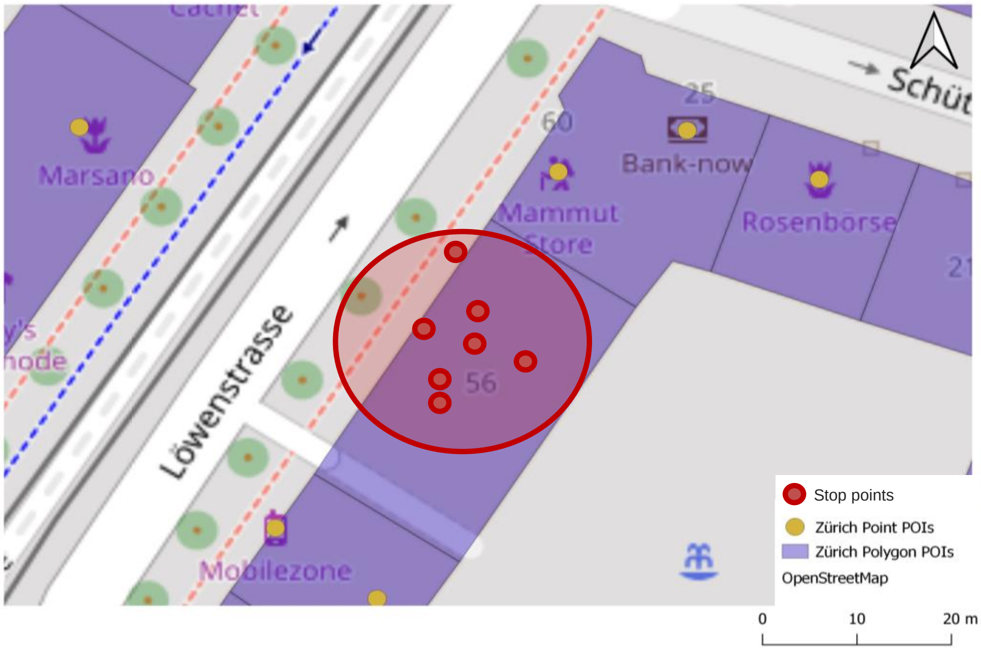

Figure 17 A possible stop episode with most of the stops inside the building

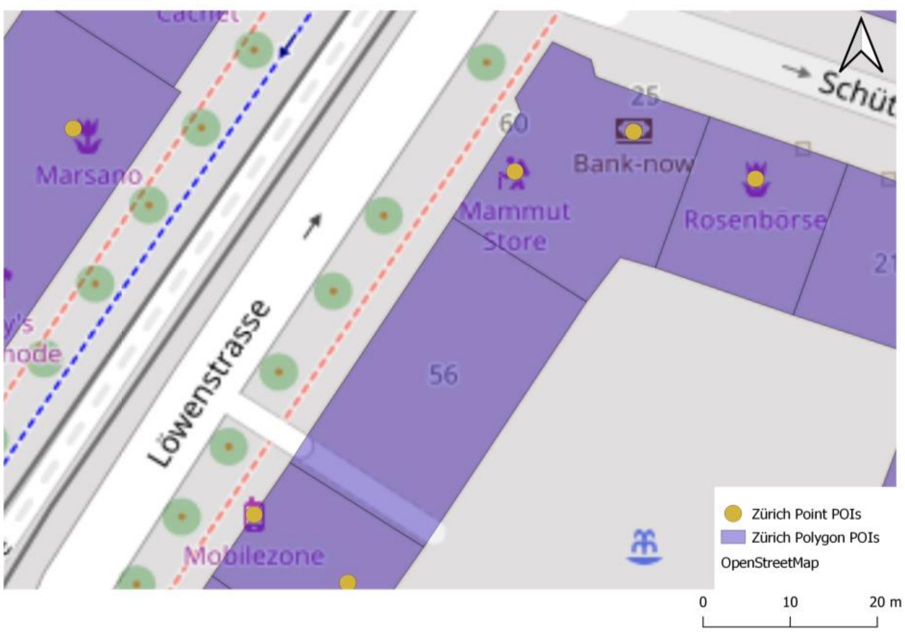

Figure 18 A sample image of the polygon POIs and point POIs. Some of the polygons include multiple points and some include none. The polygons are not associated with the information about the points. This makes it challenging to associate the stop episodes to the correct feature.

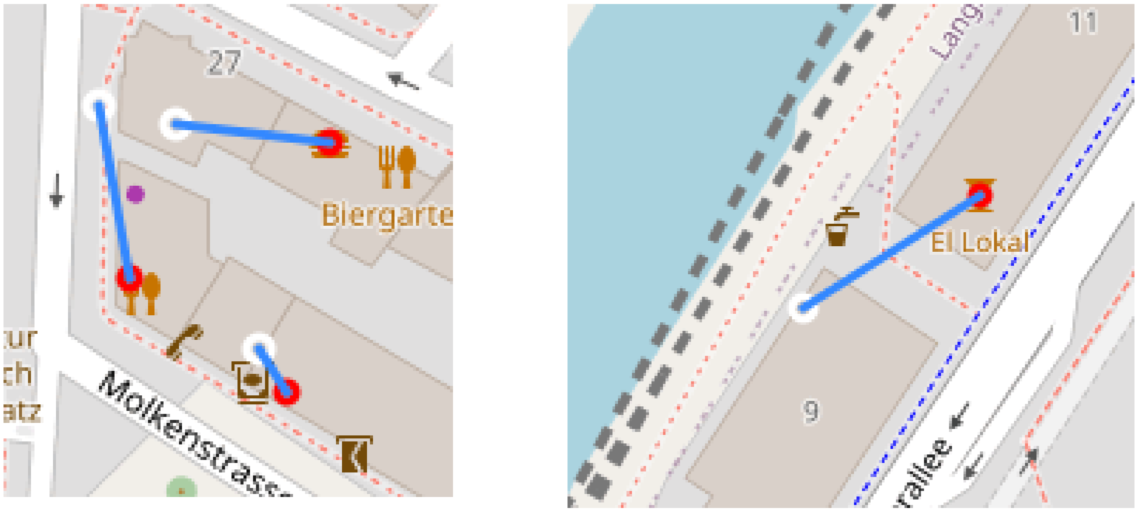

Figure 19 Possible stop episodes where the centroid of the stop episode is illustrated in white and the associated POI in red. When simply taking the nearest neighbour, this can lead to issues as taking POI from another building or considering only the centroid which can distort the shape of the stop

Taken that there were some irrelevant place semantics, there were also semantics that were more relevant than others for our purposes. For example, a café would be relevant, whereas the information that the café has a toilet was not. Therefore, we set a semantic hierarchy to define the order of the semantics that are meaningful, which was then used to extract the correct tags from OSM data. To derive the most relevant tags, we used a similar method to Bermingham and Lee (2019), where they added weights to the tags as a descriptiveness score. We set a similar hierarchy to theirs and overwrite the previous tags if tags with a higher hierarchy level tag was present. To see the full hierarchy of the semantics, please refer to Appendix section C.

The enrichment of semantics of the stop episodes with POIs was done by using a mixed model, which takes into account the distance and the duration of the nearest neighbor (as done by Ruta et al. (2012)). This assigns spatial semantics, e.g., name and social function of a place, to the stop episodes. There are also many models that use deep learning to connect the most likely POI to the stop episode. However, these models incorporate other information to the model as well, such as the other POIs visited by the same person or by other people, and this would create an inherent bias towards certain POI types and visitation patterns. As we compared these precise visitation patterns of POI types between different people, if the data were biased in this way, it would have carried on to our final conclusions. Thus, we used a more physical model.

Our adapted model iterates over each stop episode to calculate the most likely POI associated with it based on a mixed model with a hierarchical decision tree. First, we calculated all the distances of stop points to each nearest point POI. We calculated the intersection of both the land use polygons and normal polygons for each stop point. We then assigned each stop episode to three POIs (point, polygon, and land use polygon) by calculating the most common POI from POIs associated to stop points. We then compared the different POIs by comparing the median distances to point POIs, the counts of the POIs, and the sizes of the polygon POIs, in case of an intersection of polygon POIs. This is done with a hierarchical decision tree where the POIs need to have a certain amount of associated stop points within a certain distance to be selected. The full code and the thresholds associated with the different POI types are shown in the Appendix section G with the pseudo code of the semantic enrichment. A sample trajectory of the stop episodes with the associated POIs is shown in Figure 20 where 'landuse residential' is a land use polygon POI and 'shop garden_centre' is a point POI.

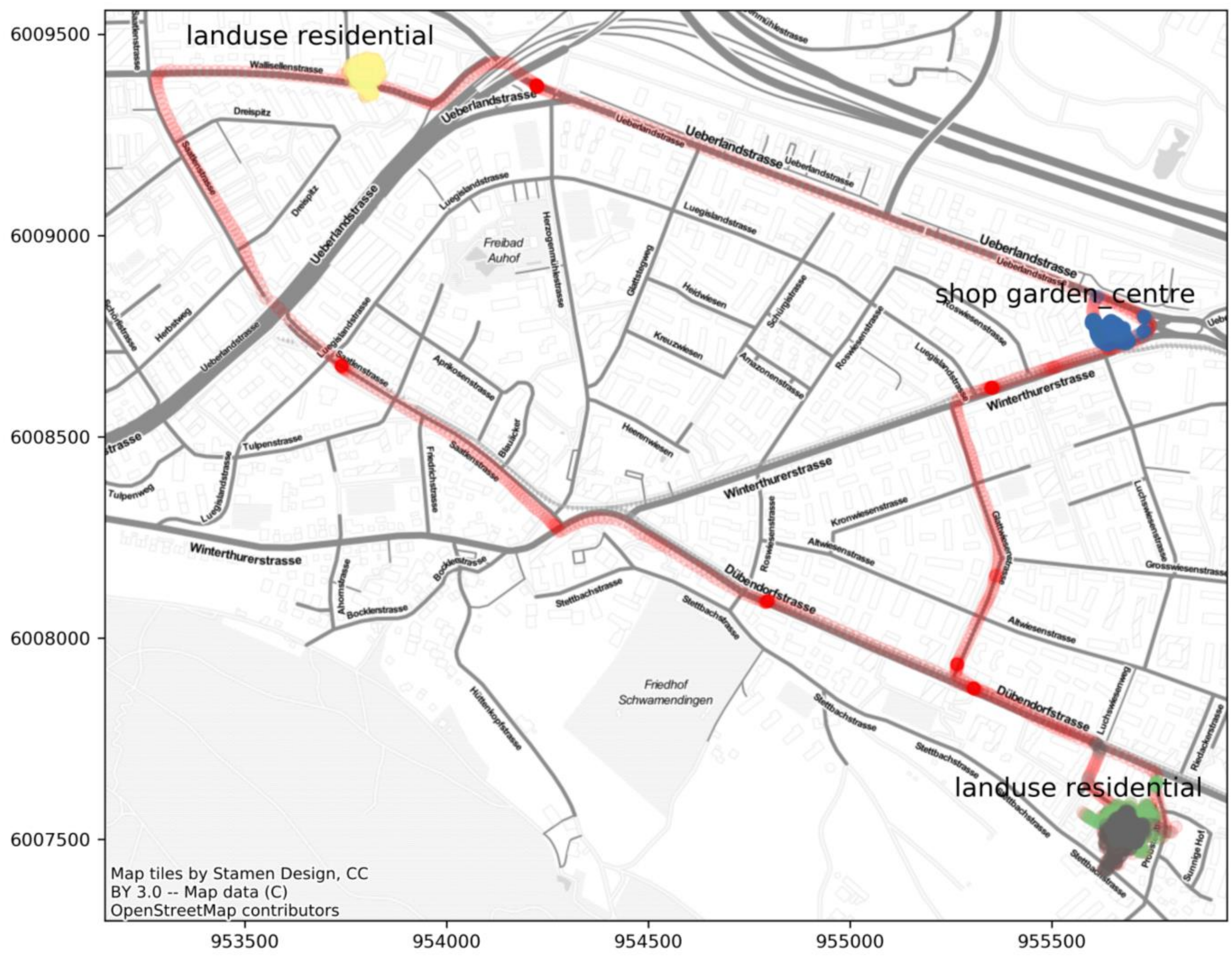

Figure 20 A sample result of semantic enrichment

5.5 Place modeling and clustering with the embeddings

We modeled the place similarities through the sequential pattern of stop episodes (Figure 9c and d). Traj2Vec enables modeling of the types of places through their position in the trajectory by adapting a machine learning technique from a word-based system. We then clustered the outputs of the modeled places using hierarchical clustering. We used complete-linkage clustering, which initially treats each observation as a separate cluster and then iteratively combines the observations together based on the distance, for which cosine similarity was used. We used all the days that had a minimum of 6h of data, which resulted in 124 participants that had at least one day included.

Traj2Vec (Zhang et al., 2020) is based on Word2Vec, which models the word similarities through textual input and the context in which the word appears. Word2Vec models the context related to the word, so for example, in a case of synonyms, their context would often be similar. Therefore they would also be very closely modeled in the vector space. In a trajectory, this would mean, for example, that a person is going to a store, but the type of store does not matter. Therefore a corner store, a supermarket, and a grocery store would get similar vectors. The assumption here lies that people would have similar types of trajectories for similar functions. This could also help to derive the functions that are similar in affordance but might not be similar in the 'common sense' human semantics. For example, if people would use forest or a library in the same way, e.g., they can both be relaxing, and thus similar in their relationship with health, although forest and library are very different in their 'common' semantics. Word2Vec takes sentences as input to create high dimensional vectors for each word, whereas Traj2Vec takes the sequences of places as an input and creates a vector for each place. We tested aggregating places by days as sentences vs. persons as sentences. We aimed to model the frequent semantics across our sample of different place similarities based on their occurrence in the trajectories with Traj2Vec.

For the place embedding with Traj2Vec, we used a minimum count of five, as these POIs could be outliers, with 20 dimensions, window size of five, and Continuous Bag of Words (CBOW) model. CBOW uses the surroundings to predict the target object, in comparison to Skip-gram, which predicts the environment rather than the target. This is further illustrated with place names in Figure 21, where the 'pharmacist' has been replaced with a 'chemist shop' in the semantic trajectory that provides essentially the same function and the environment. CBOW is more accurate with frequent elements rather than infrequent ones, and as we are modeling the frequency of the place visitations, the very infrequently visited places are not of high importance for this analysis. For the clustering, we used hierarchical clustering as it does not require dimensional reduction which would lose information and works well with medium-sized data. For the hierarchical clustering, we used complete linkage with the cosine similarity as Euclidean distance loses its meaning in high dimensions. The illustrations of the dendrograms were created with 'scipy cluster hierarchy' python module, and the cluster tags were then derived by using 'sklearn cluster' module.

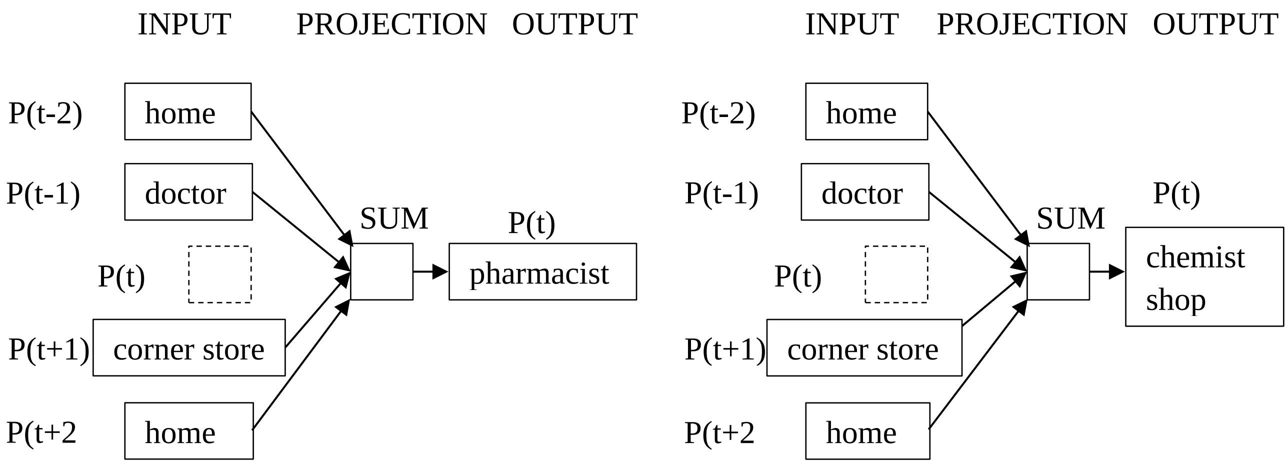

Figure 21 Illustration of CBOW with place names

After the place embedding, the hierarchical clustering aggregated the visited POI types to groups according to their similarity and relatedness using the outputs from Traj2Vec. This allowed the examination of the relationship between different clusters and health. Rather than classifying the places as specific places with only the tag names from OSM, the places were classified according to their spatial context within a trajectory and the function. The similarities of the place vectors were then measured by using cosine vector space similarities. This added a lower granularity dimension to the data that could be analyzed statistically.

The relevance of moving between granularities can be illustrated, for example, when considering 'the University Hospital in Zurich,' which is one group of buildings with an exact address, whereas 'a hospital in Zurich' has lower granularity. Different granularities are used for different purposes, and sometimes it is not necessary to define which hospital, but the knowledge of a hospital would be sufficient. Another way would be to model the relaxing places or the stressful places together. To move between granularity levels, specialization or generalization is required (Fonseca et al., 2002). This study is concerned about generalization, and thus, we added lower levels of granularity to the visited places. Using vectors extracted from the trajectories' sequences, we were able to move to a lower granularity level by clustering the places together that are similar in their surrounding trajectories.

The importance of clustering place types is highlighted when there are many places that have only very few visitations. Often some of the POIs are visited very often, whereas some very rarely. This can lead to problems when testing for correlation in skewed data. Therefore, moving to a lower granularity level can combine some of the more rarely visited POIs together and thus reduce some of the skewness which makes the statistical analysis possible.

We also adapted the Mot2vec model from Crivellari and Beinat (2019), who used it to visually analyze the vector spaces between the users. Mot2vec averages all the place vectors that a participant has visited and thus creates each participant a vector, which can be used to compare the individuals between each other. Their method also includes the visitation times by splitting the days into time intervals, which are then associated with a place. We used the sequences of POIs rather than timestamped sequences as this reduces the possible errors in the timing of the start and end times of the stop episodes as well as allows us to compare the outputs from the Traj2Vec place modeling based on the same place vectors. Therefore, for each participant, from the sample that we used for our statistical analysis, we averaged the visited place vectors for each participant and visualized them with t-SNE with different hues based on the health indicator values.

To compare the modeled clusters of Traj2Vec, we also modeled the clusters based on 'common sense.' These are clusters that a person would associate together based on their label similarities intuitively. To find an exhaustive list of the POI allocation into these clusters, please see the Appendix Section C. The modelling tests, whether this type of clustering is meaningful to the semantic place modeling.

5.6 Multivariate regressions between the place visitations and health indicators

Multivariate regression models (Figure 9e) were conducted between the different types of places and health indicators to investigate whether they are related to health through frequency. For all the models, we tested the skewness of the variables, and if the variables were overly skewed, they were excluded from the models. However, sometimes the skewness of the variable could be reduced to an acceptable level by using log or cube transformation. Thereafter, we used forward and backward selection to select the most important variables. We included age and gender to all of the models as control variables, although in some cases, they were excluded during the forward and backward selection. All the models were conducted with the statistically valid sample of six random days from each participant. We used HC1 robust standard errors to minimize the effect of the outliers.

We conducted the statistical analysis through the following steps: First, we compared the individual POI types and their relation to the health indicators to see whether there are specific place types with a relationship to health. This also acted as a baseline for the cluster models. We tested for the skewness of the POI types to see if the POI types can be used for the models. Second, we regressed the variety of the different POI types and the total count of POI types visited. To estimate variety, we took the count of all the unique POI types for each participant, whereas the total count was measured by the number of all of the visited POI types per participant. These measures were then used in a multivariate regression with the health indicators. Third, we compared the correlations between different levels of hierarchical clusters and health indicators. We moved from lower levels of clusters to higher levels to investigate whether certain clusters have a particular relationship with health. We took POI cluster counts of 30, 20, 14, and 8, as these seem to have small breakpoints. We then compared the POI cluster counts of the participants to the health indicators. Lastly, we compared our results to 'common sense' cluster regression. The common sense clusters were created to mimic semantic associations that humans intuitively have of place types. This intuitive taxonomy of clusters can be seen in Appendix Section C. We used 'key' as a primary indicator of the semantic meaning as well as their common function. Here we also calculated the POI cluster counts for each person and compared the frequencies of each cluster with the health indicators.

6 Results

6.1 Preprocessing results: Trajectory segmentation and enrichment

In the trajectory segmentation, we separated the movements from the stops. Figure 22 illustrates that stops generally have very low speeds and movements have a larger range of speeds. There are also some low speeds in the movement, which emphasizes that speed-based parameters have problems with detecting low speed stops when a person is walking, for example. We can also see some noise in the stop detection as some stop observations have relatively high speeds. However, this is unlikely to cause any issues due to the nature of our semantic enrichment which is not sensitive to single observations.

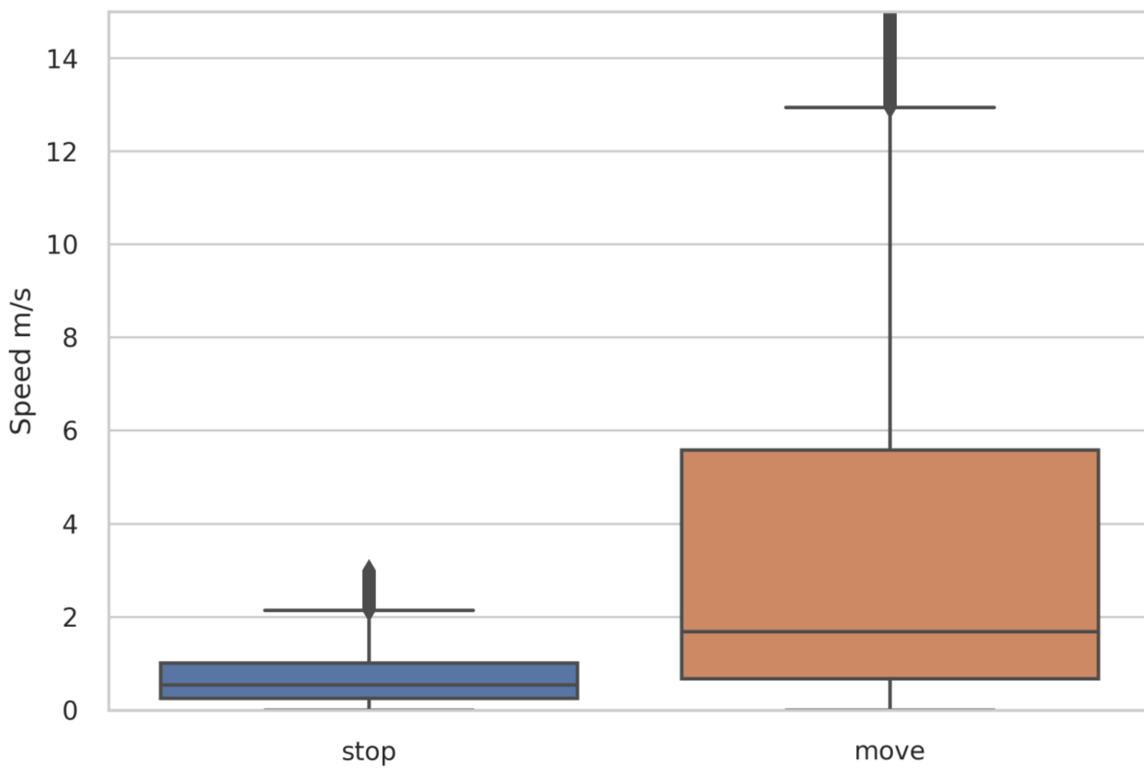

Figure 22 The frequencies of speeds between stops and moves. The image is capped to 15m/s, however, there are scattered speeds of up to 70m/s in the moves.

Figure 23 shows that most of the stop episodes are from the Zurich region, which is also where most of the GPS way point observations are from.

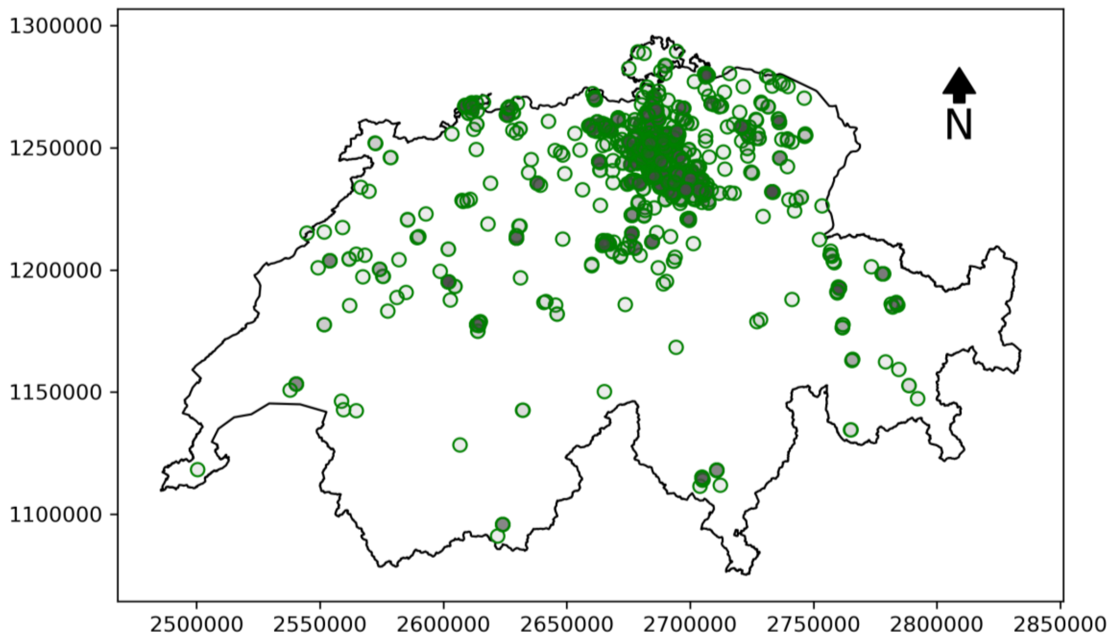

Figure 23 All the centroids of the stop episodes

Traj2Vec creates place embeddings based on sentences. The hierarchical clustering was used to examine the results of aggregation of sentences by days vs. persons. Aggregating place sequences by days resulted in very poor skewness (as can be seen in the histograms in Figure 24), possibly due to the nature of the data. Aggregating by persons as sentences reduced the skewness (Figure 25). Figure 24 and Figure 25 show that towards lower granularity, i.e., towards the right side of the image with smaller clusters, the skewness generally increases as there are fewer observations in general.

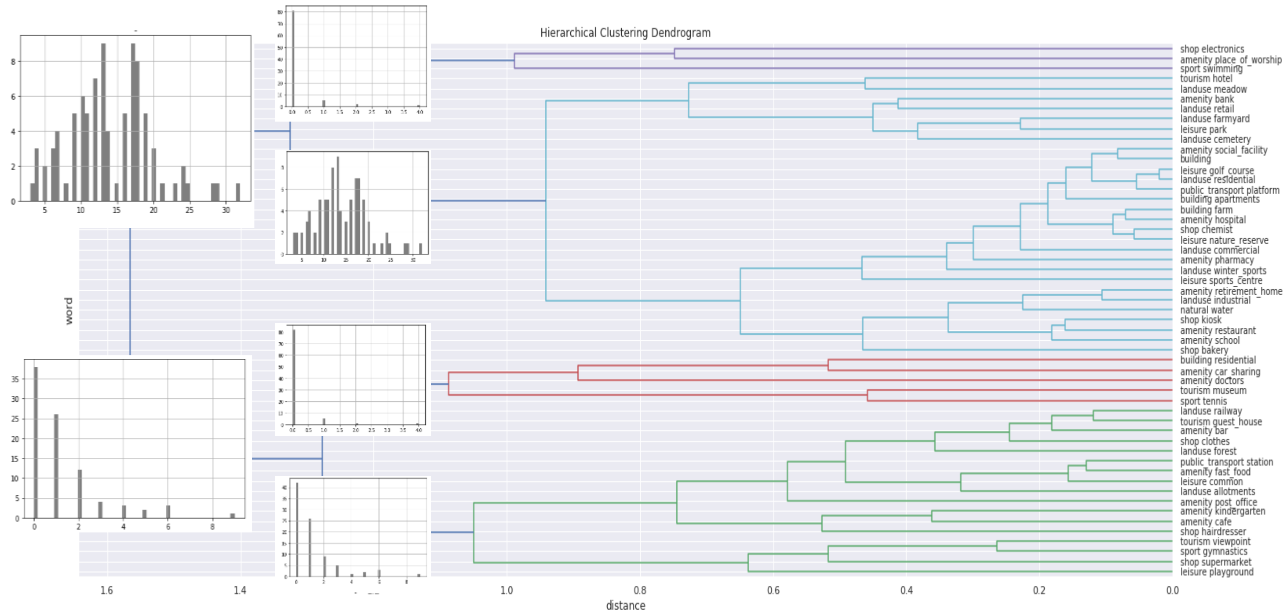

Figure 24 Dendrogram of the hierarchical clusters using aggregated places of days as sentences. Histograms show the distribution of visitation frequencies

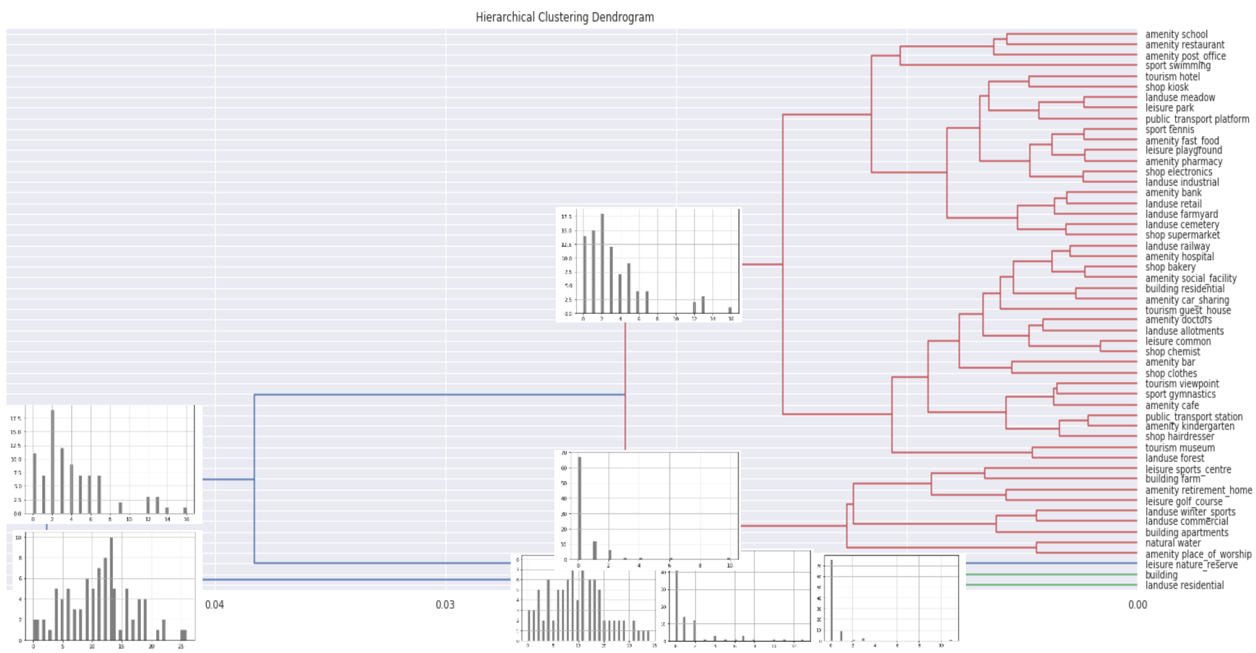

Figure 25 Dendrogram of the hierarchical clustering using aggregated places of persons as sentences. Histograms show the distribution of visitation frequencies

6.2 Comparison between the statistical sample and participants without sufficient GPS data

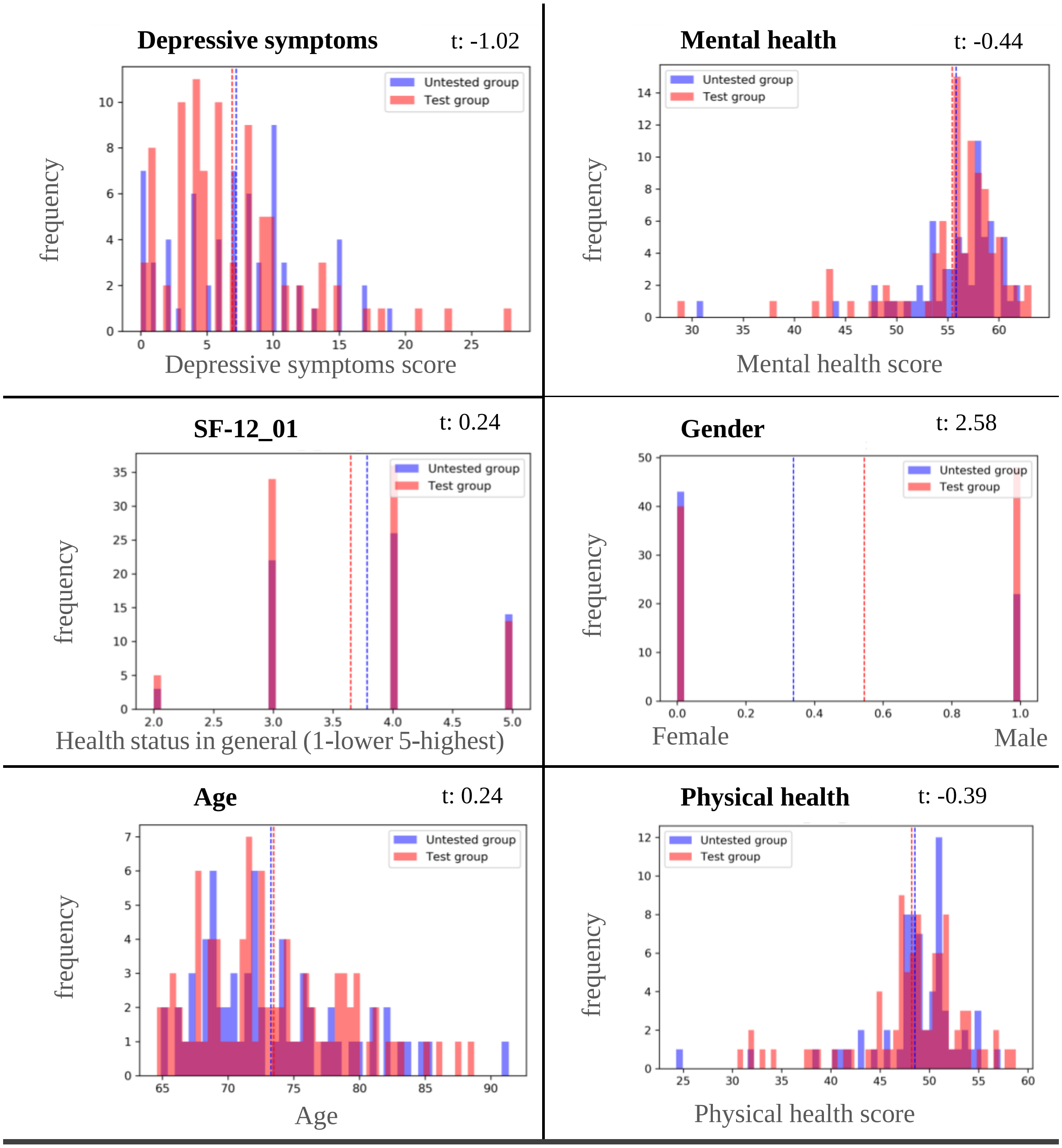

Figure 26 Comparison of histograms between participants with sufficient data and insufficient GPS data available

Only a part of the data was incorporated into the statistical analysis to ensure that all the participants in the statistical analysis have enough GPS data. We took random six days from days that had a minimum of 6h availability from 88 people. This left 65 people who did not have sufficient GPS data for the statistical sample. To make sure that there was no difference between the people that had more GPS data, we tested for the different health indicators and whether they were significantly different. The histograms and the t-values are illustrated in Figure 26. Only gender showed a statistical difference in our groups, which could be due to the relatively small sample size. However, as shown later sections, gender has very little impact on our models.

6.3 The relationship between individual POI types and health

The frequencies in the POI visitations are highly skewed (Figure 27). These frequencies are from the days that had a minimum of six hours of data per day and were used for Traj2Vec model training. 124 participants were thus included as these had at least one day that fulfilled the criteria.

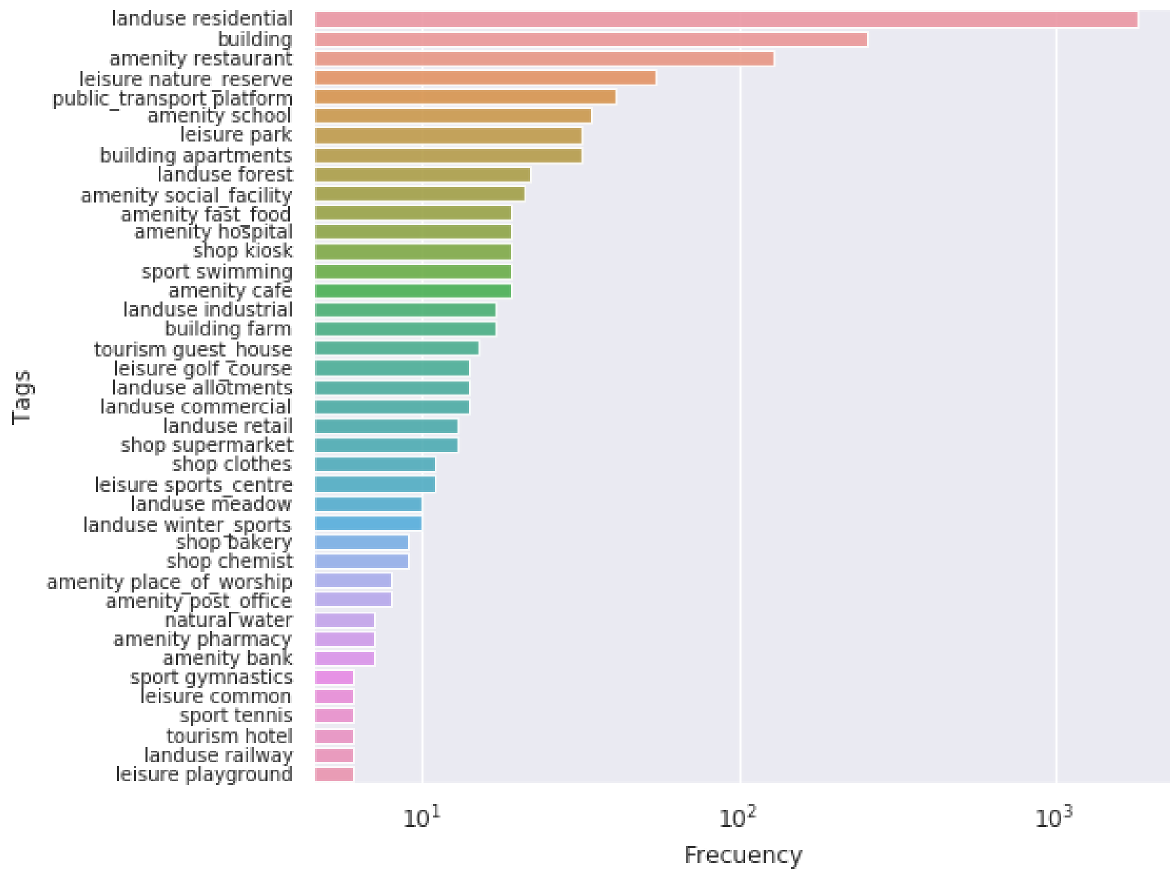

Figure 27 Frequencies of the place visitations. The visitation counts are highly skewed, and the frequency is logarithmic to allow some comparison also of the lower values. The minimum count is five which is also used for the Traj2Vec training.

The individual POI types are highly affected by skewness. Therefore only nine POI types could be selected for the regression model even after log or cube transformation. After forward and backward selection six POI types were included in the models, as can be seen from Figure 28. The other three that were not included are: ‘building’, ‘landuse residential’, and ‘leisure park’. The results show some correlations between the health indicators and POI type visitation. ‘Forest’ and ‘nature’ are both associated with higher mental health and ‘nature’ with lower depression. ‘Fastfood’ is associated with lower mental and physical health, whereas a ‘café’ with higher mental health. The correlations are generally very weak as R2 is especially small in other models other than in the mental health model.

	Dependent variable:			
	bl_sf12_mental	bl_h_SF12_01	bl_sf12_physical	blADS
	(1)	(2)	(3)	(4)
amenity fastfood (cube)	-2.815	-0.390*		
	(2.617)	(0.236)		
landuse forest (cube)	3.253***			
	(0.947)			
leisure nature(cube)	2.316**			-1.788*
	(1.006)			(0.923)
amenity restaurant (log)	-3.278			
	(2.381)			
amenity cafe	2.079**			
	(1.060)			
shop supermarket				3.250
				(3.046)
age			-0.280**	0.182***
			(0.130)	(0.067)
blsex			1.916	
			(1.199)	
Constant	55.766***	3.686***	67.741***	-6.522
	(0.786)	(0.090)	(9.258)	(5.016)
Observations	88	88	88	88
R2	0.176	0.023	0.092	0.094
Adjusted R2	0.126	0.011	0.071	0.061
Residual Std. Error	5.168 (df = 82)	0.798 (df = 86)	5.377 (df = 85)	5.049 (df = 84)
F Statistic	3.504*** (df = 5; 82)	1.995 (df = 1; 86)	4.300** (df = 2; 85)	2.891** (df = 3; 84)
Note:				*p<0.1; **p<0.05; ***p<0.01

Figure 28 Regression results of the individual POI types.

6.4 The number of unique and total visitations of POI types and the relation to health

We tested the correlations between having a higher number of unique POI types, and the number of total POI types visited within the six day period. As after forward and backward selection neither of these variables showed any correlation with health, Figure 29 shows all the coefficients, which are all unsignificant after controlling for age and gender.

	Dependent variable:			
	bl_sf12_mental	bl_sf12_physical	bl_h_SF12_01	blADS
	(1)	(2)	(3)	(4)
uniquePOIs	0.609	-0.521	-0.058	-0.061
	(1.108)	(0.754)	(0.100)	(0.710)
total	-0.035	0.032	0.035	-0.033
	(0.311)	(0.196)	(0.024)	(0.176)
Constant	51.391***	69.140***	4.760***	-5.586
	(9.221)	(9.823)	(1.018)	(5.526)
Age Effects	Yes	Yes	Yes	Yes
Gender Effects	Yes	Yes	Yes	Yes
Observations	88	88	88	88
R2	0.027	0.103	0.049	0.051
Adjusted R2	-0.019	0.060	0.003	0.005
Residual Std. Error (df = 83)	5.581	5.407	0.802	5.198
F Statistic (df = 4; 83)	0.586	2.389*	1.072	1.107
Note:		*p<0.1; **p<0.05; ***p<0.01		

Figure 29 Regression results of the variety and the total number of POI visitations.

6.5 The relationship between the hierarchical clusters of POI types from Traj2Vec and health

6.5.1 30-Clusters of POI types scenario

Figure 30 shows the cutoff point for the clusters. With a high number of clusters, we aim to cluster together only the places that are very similar. After forward and backward selection, eight clusters have been selected for the regression models. The clusters that have been selected for the models based on relatively low skewness can be seen in the Appendix section D. We can see from Figure 31 that many of the clusters combine together places that we would not normally assign in the same clusters. This illustrates the idea that places can have similarity in a trajectory without being necessarily similar in the names of the places. Table 1 shows the health indicators associated with the clusters. Appendix Section E shows all the clusters and their POI types. We can see that R2 has improved in the SF12_01 and ADS in comparison to the individual POI-model.

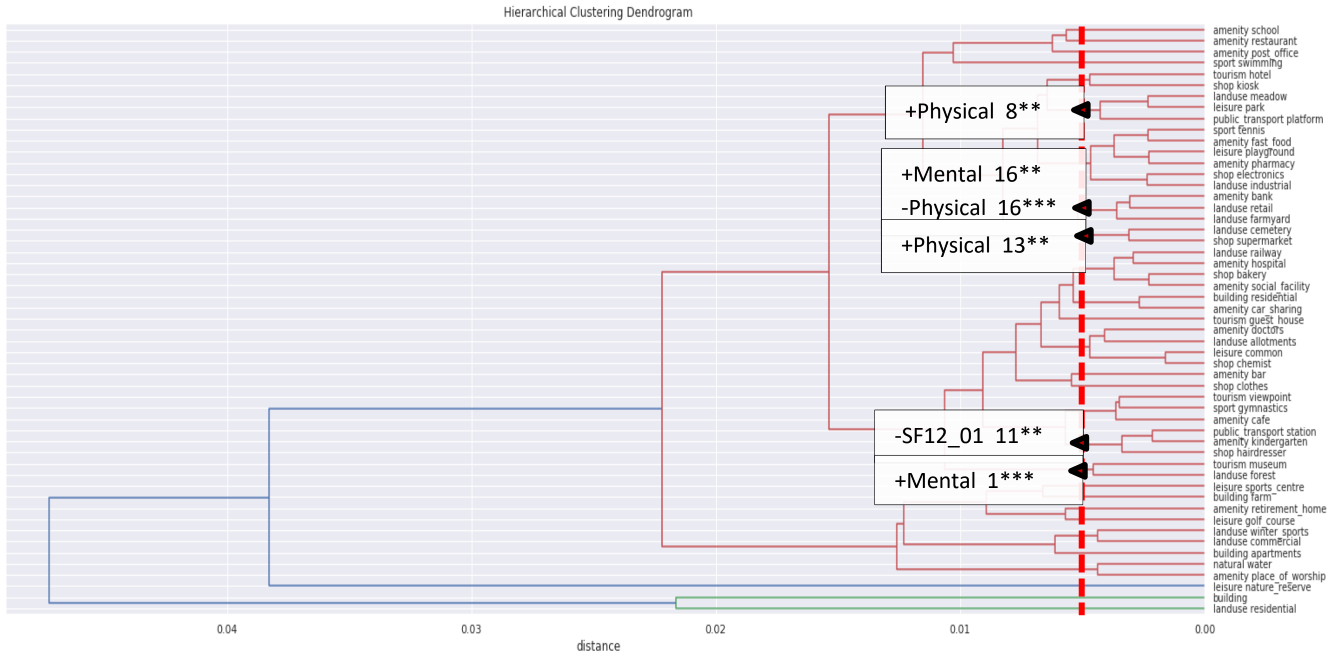

Figure 30 Cutoff for 30-clusters at the red dashed line, where the POI types that are joined on the right side of the line are combined in a cluster.

Table 1 30-cluster model's clusters selected to the multivariate models and the POI types of the clusters

Cluster 1 +Mental ***	Cluster 8 +Physical **	Cluster 11 -SF12_01 **	Cluster 5	Cluster 16 +Mental ** -Physical ***	Cluster 13 +Physical **
landuse forest tourism museum	public_transport platform leisure park landuse meadow	amenity kindergarten shop hairdresser public_transport station	amenity social_facility amenity hospital shop bakery landuse railway	amenity bank landuse farmyard landuse retail	shop supermarket landuse cemetery

.

	Dependent variable:			
	bl_sf12_mental (1)	bl_sf12_physical (2)	bl_h_SF12_01 (3)	blADS (4)
1 (cube)	3.007*** (0.892)			
8 (log)		2.871** (1.214)		
11 (cube)			-0.581** (0.280)	3.230 (2.337)
5			-0.304 (0.198)	1.922 (1.574)
16	2.640** (1.345)	-3.046*** (1.155)		
13		2.636** (1.221)	-0.398 (0.268)	3.414 (2.349)
landuse residential		-0.156 (0.100)		
leisure nature(cube)	2.775*** (1.032)			-1.686** (0.839)
amenity restaurant (log)	-3.305 (2.265)			
age		-0.265** (0.123)		0.173** (0.068)
blsex		2.173* (1.233)		
Constant	55.413*** (0.900)	67.260*** (8.962)	3.805*** (0.100)	-6.614 (5.058)
Observations	88	88	88	88
R2	0.155	0.175	0.086	0.152
Adjusted R2	0.115	0.114	0.053	0.100
Residual Std. Error	5.202 (df = 83)	5.249 (df = 81)	0.781 (df = 84)	4.943 (df = 82)
F Statistic	3.814*** (df = 4; 83)	2.870** (df = 6; 81)	2.628* (df = 3; 84)	2.931** (df = 5; 82)

Figure 31 Regression results of the 30-clusters

6.5.2 20-Clusters of POI types scenario

Figure 32 shows the cutoff for the 20-clusters where the cutoff has been now moved towards the left, and more POI types are joined together on the right side. Some of the previous clusters from the 30-clusters are also present in the current regression models after filtering for skewness and backward and forward selection. Clusters $13^{30\text{-cluster}}$ and $16^{30\text{-cluster}}$ have now been combined into Cluster $14^{20\text{-cluster}}$ that does not have a significant relationship with the health indicators, whereas $13^{30\text{-cluster}}$ had a positive and $16^{30\text{-cluster}}$ negative association with physical health. Cluster $8^{30\text{-cluster}}$ remains as $8^{20\text{-cluster}}$ and has a significant positive relationship with physical health, as well as Cluster $1^{30\text{-cluster}}$ is equal to Cluster $4^{20\text{-cluster}}$ and has the same positive relationship with mental health. Cluster $11^{30\text{-cluster}}$, is dropped from the models. Cluster 5[30-cluster], which was previously not significant in the 30-cluster scenario, has been combined with another cluster and is significant with a negative association with physical health.

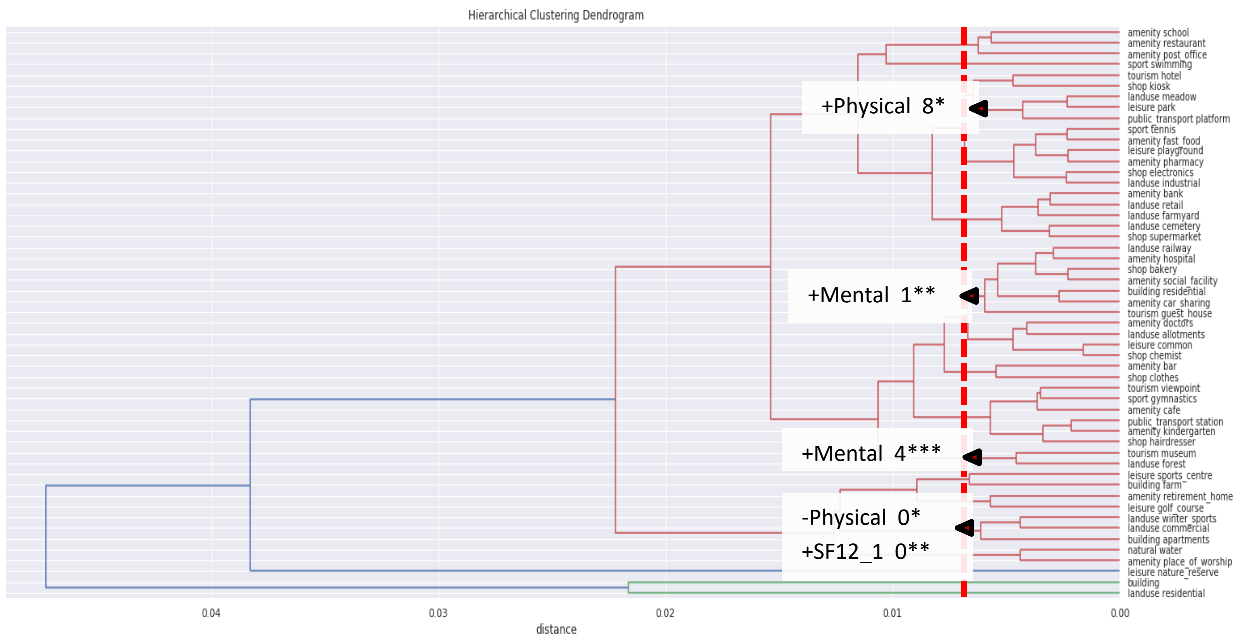

Figure 32 Cutoff for 20-clusters at the red dashed line, where the POI types that are joined on the right side of the line are combined in a cluster.

Table 2 20-cluster model's clusters selected to the multivariate models and the POI types of the clusters

Cluster 4 +Mental ***	Cluster 3	Cluster 0 -Physical *, +SF12_1 **	Cluster 1 -Physical **	Cluster 8 +Physical *	Cluster 14
landuse forest tourism museum	amenity restaurant amenity school	building apartments landuse commercial landuse winter_sports	amenity social_facility amenity hospital tourism guest_house shop bakery landuse railway amenity car_sharing	public_transport platform leisure park landuse meadow	shop supermarket landuse retail amenity bank landuse cemetery landuse farmyard

	Dependent variable:			
	bl_sf12_mental	bl_h_SF12_01	bl_sf12_physical	blADS
	(1)	(2)	(3)	(4)
3 (log)	-2.825			
	(2.078)			
4 (cube)	2.662***			
	(0.831)			
0 (log)		0.487**	-2.612*	
		(0.240)	(1.577)	
1 (cube)			-3.267**	
			(1.390)	
8 (log)			1.826*	
			(1.012)	
14				2.032
				(1.969)
landuse residential			-0.180**	
			(0.088)	
leisure nature(cube)	2.386**			
	(0.998)			
age			-0.220*	0.185***
			(0.130)	(0.068)
Constant	55.806***	3.580***	66.633***	-7.216
	(0.869)	(0.087)	(9.500)	(5.126)
Observations	88	88	88	88
R2	0.115	0.048	0.195	0.067
Adjusted R2	0.083	0.037	0.145	0.045
Residual Std. Error	5.293 (df = 84)	0.788 (df = 86)	5.156 (df = 82)	5.091 (df = 85)
F Statistic	3.629** (df = 3; 84)	4.380** (df = 1; 86)	3.961*** (df = 5; 82)	3.069* (df = 2; 85)
Note:				*p<0.1; **p<0.05; ***p<0.01

Figure 33 Regression results of the 20 clusters.

6.5.3 14 Clusters of POI types scenario

The 14-cluster scenario's clustering cutoff, table for the allocation of POI types in the clusters, and the correlation table are shown beneath. The only significant explanatory variables are Cluster $0^{14\text{-cluster}}$ and Cluster $8^{14\text{-cluster}}$. Cluster $8^{14\text{-cluster}}$ is the same as Cluster $0^{20\text{-cluster}}$. Cluster $0^{14\text{-cluster}}$ encompasses a large number of POI types. It includes POI types, such as hospital and social facilities that have been negatively associated with health in previous models as well. It also includes other health-related POI types such as chemist and doctors, which were previously not included due to too high skewness. 'shop clothes' and 'amenity bar' make another cluster that had also not been previously included due to skewness but is now included in the Cluster $0^{14\text{-cluster}}$ as well.

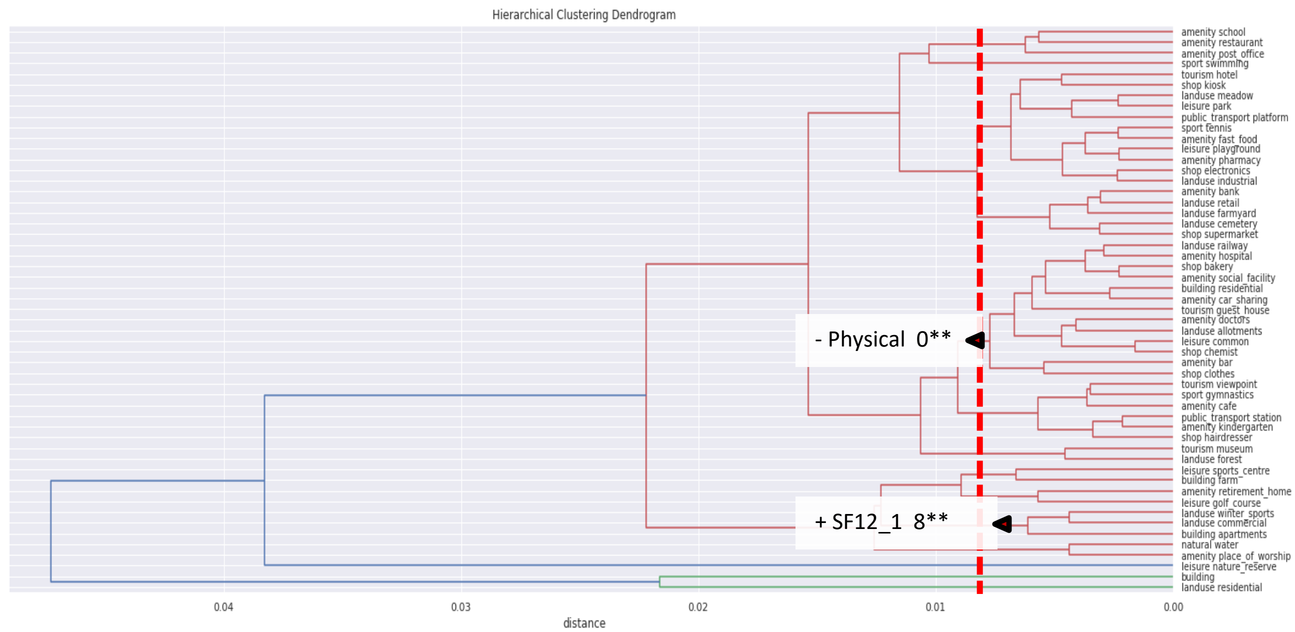

Figure 34 Cutoff for 14 clusters

Table 3 14-cluster model's clusters selected to the multivariate models and the POI types of the clusters

Cluster 2	Cluster 0 - Physical **	Cluster 6	Cluster 1	Cluster 8 + SF12_1 **
amenity restaurant	amenity social_facility	shop supermarket	public_transport platform	building apartments
amenity school	amenity hospital	landuse retail	leisure park	landuse commercial
amenity post_office	tourism guest_house	amenity bank	shop kiosk	landuse winter_sports
	landuse allotments	landuse cemetery	amenity fast_food	
	shop clothes	landuse farmyard	landuse industrial	
	shop bakery		landuse meadow	
	shop chemist		amenity pharmacy	
	landuse railway		leisure playground	
	leisure common		tourism hotel	
	amenity bar		sport tennis	
	amenity car_sharing			
	building residential			
	amenity doctors			

	Dependent variable:			
	bl_sf12_mental (1)	bl_sf12_physical (2)	bl_h_SF12_01 (3)	blADS (4)
2 (log)	-2.454 (2.042)			
0 (log)		-2.643** (1.143)		
6				2.032 (1.969)
1 (log)			0.214 (0.136)	
8 (log)			0.514** (0.225)	
landuse residential		-0.188** (0.088)		
leisure nature(cube)		-2.197 (1.455)		
age		-0.294** (0.131)		0.185*** (0.068)
blsex		1.801 (1.138)		
Constant	56.486*** (0.834)	71.675*** (9.544)	3.479*** (0.110)	-7.216 (5.126)
Observations	88	88	88	88
R2	0.058	0.195	0.074	0.067
Adjusted R2	0.047	0.146	0.052	0.045
Residual Std. Error	5.395 (df = 86)	5.154 (df = 82)	0.782 (df = 85)	5.091 (df = 85)
F Statistic	5.325** (df = 1; 86)	3.977*** (df = 5; 82)	3.373** (df = 2; 85)	3.069* (df = 2; 85)
Note:				*p<0.1; **p<0.05; ***p<0.01

Figure 35 Regression results of the 14 clusters

6.5.4 8 Clusters of POI types scenario

The 8-cluster scenario divides the POI types into two larger groups, three small groups, and three single POI types. We can see that Cluster 3 is the same as Cluster $8^{14\text{-cluster}}$ and Cluster $0^{20\text{-cluster}}$. The Cluster $2^{8\text{-cluster}}$ combines Cluster $0^{14\text{-cluster}}$, Clusters $1^{20\text{-cluster}}$ and $4^{20\text{-cluster}}$ and the Cluster $11^{30\text{-cluster}}$. Three of the four clusters correspond with the same correlations as Cluster $0^{8\text{-cluster}}$. The R2 is very low in other models than the physical health model, which is also lower than that of the 14-cluster scenario. Overall, the models show that the R2 generally decreases with higher levels of clustering, although individual POI-model had very low R2 values.

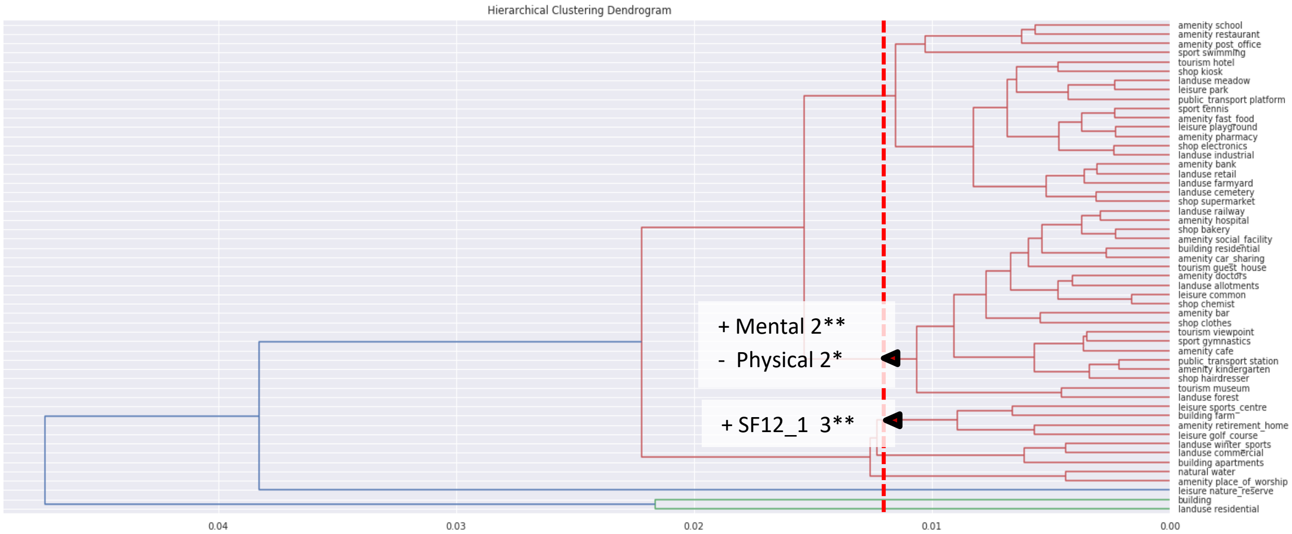

Figure 36 Cutoff for 8-clusters

Table 4 8-cluster model's clusters selected to the multivariate models and the POI types of the clusters

Cluster 2 + Mental ** - Physical *	Cluster 3 + SF12_1 **
landuse forest amenity social_facility amenity cafe amenity hospital tourism guest_house landuse allotments shop clothes shop bakery shop chemist landuse railway sport gymnastics leisure common amenity kindergarten amenity bar amenity car_sharing tourism museum tourism viewpoint building residential shop hairdresser public_transport station amenity doctors	building apartments landuse commercial landuse winter_sports

	Dependent variable:			
	bl_sf12_mental	bl_sf12_physical	bl_h_SF12_01	blADS
	(1)	(2)	(3)	(4)
2 (log)	1.625**	-1.804*		
	(0.764)	(0.964)		
3 (log)		-2.241	0.487**	
		(1.620)	(0.240)	
leisure nature(cube)	1.896*			-1.912**
	(1.017)			(0.886)
landuse residential		-0.183*		
		(0.094)		
age		-0.294**		0.160**
		(0.131)		(0.069)
blsex		1.673		
		(1.140)		
Constant	54.305***	71.881***	3.580***	-4.497
	(0.886)	(9.547)	(0.087)	(5.212)
Observations	88	88	88	88
R2	0.051	0.170	0.048	0.058
Adjusted R2	0.029	0.119	0.037	0.036
Residual Std. Error	5.447 (df = 85)	5.235 (df = 82)	0.788 (df = 86)	5.116 (df = 85)
F Statistic	2.298 (df = 2; 85)	3.348*** (df = 5; 82)	4.380** (df = 1; 86)	2.623* (df = 2; 85)
Note:				*p<0.1; **p<0.05; ***p<0.01

Figure 37 Regression results of the 8-clusters.

Table 5 shows how many POI types were selected to the models after omitting the variables for too high skewness. The number of individual POI types selected increased somewhat linearly. This shows that clustering enables the modeling of the more skewed place visitations and thus a higher variety of POI types.

Table 5 The number of POI types selected for each cluster model

Individual POI types (53)	30-clusters	20-clusters	14-clusters	8-clusters
9	30	37	46	47

6.6 The relationship between Mot2vec and health

Mot2vec was used to detect visual similarities between the participants with similar average place embeddings and the health indicators. Figure 38, Figure 39, Figure 40, and Figure 41 show that although there can be some visual clusters detected in the averaged place vectors for each participant, these do not show similarities with the health indicators. There are also no visual patterns within the clusters of the health indicators.

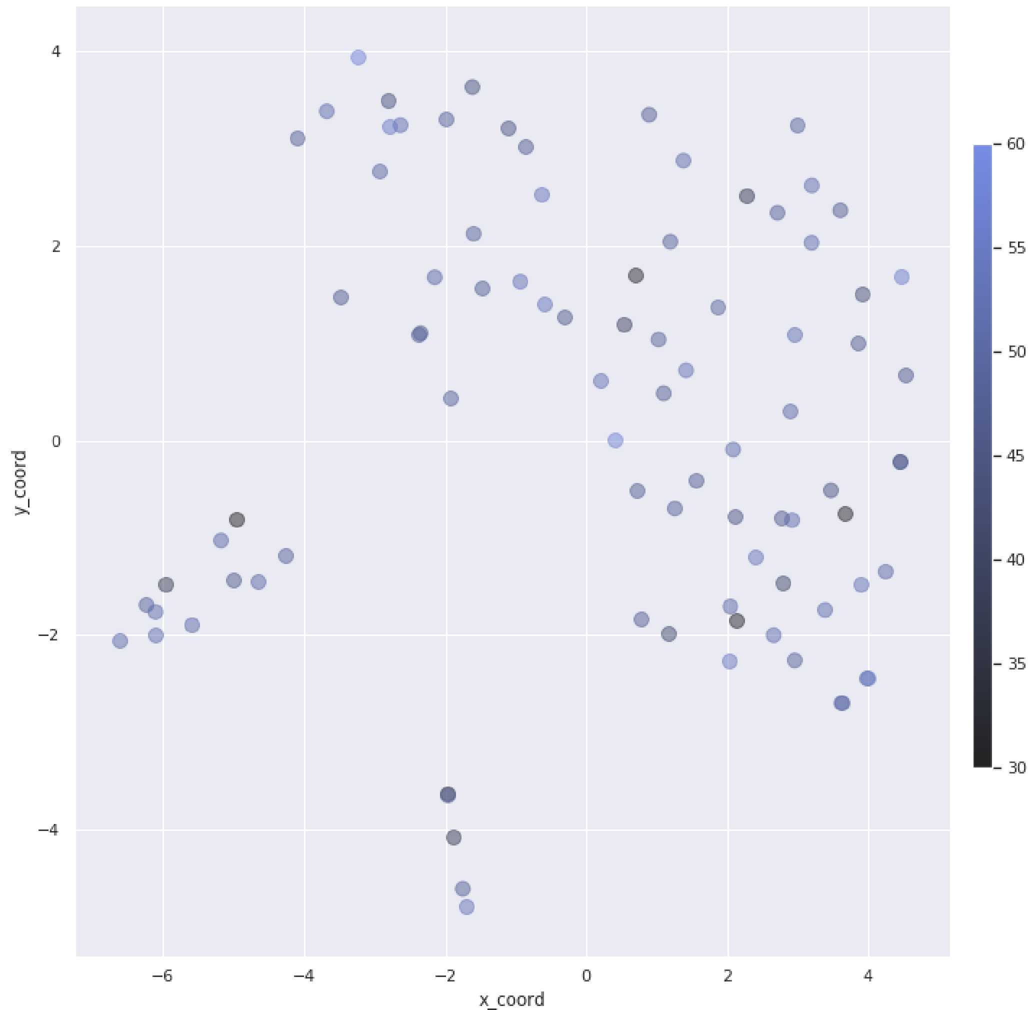

Figure 38 Averaged vectors per participant with the hue based on Physical health indicator.

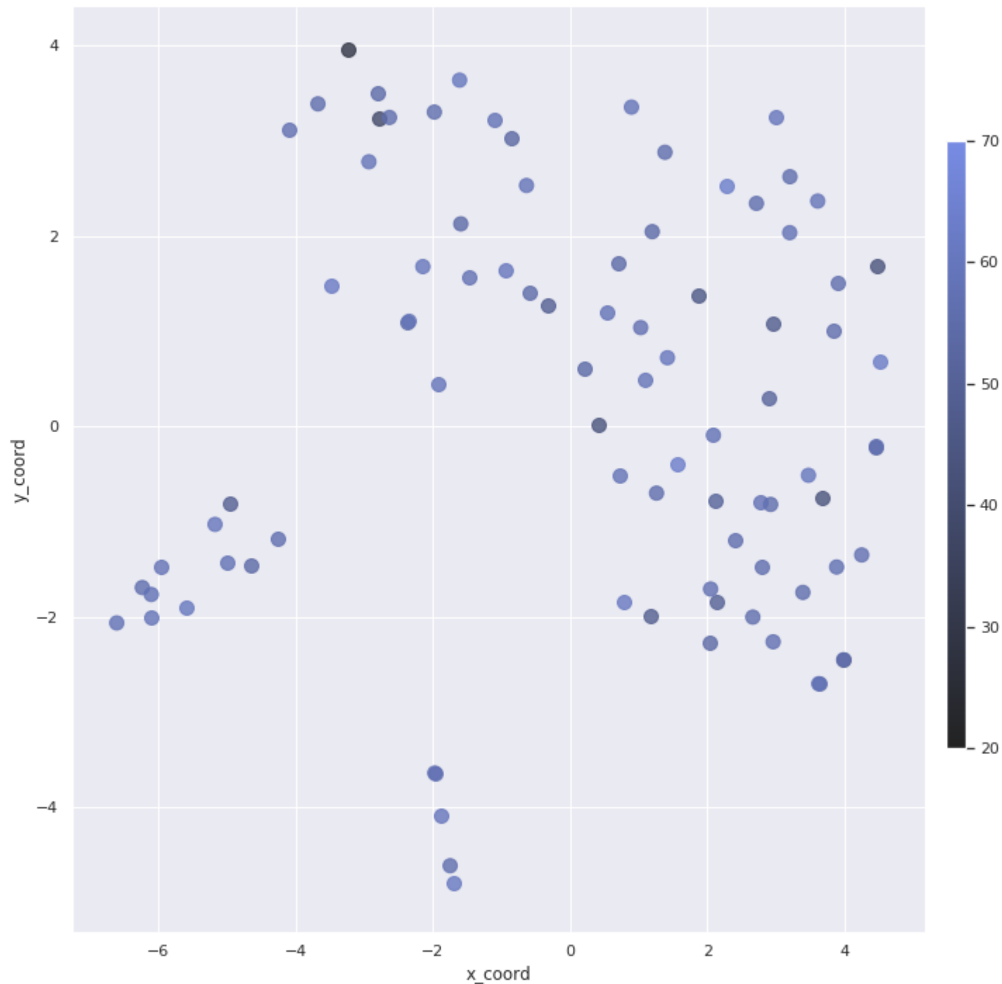

Figure 39 Averaged vectors per participant with the hue based on Mental health indicator.

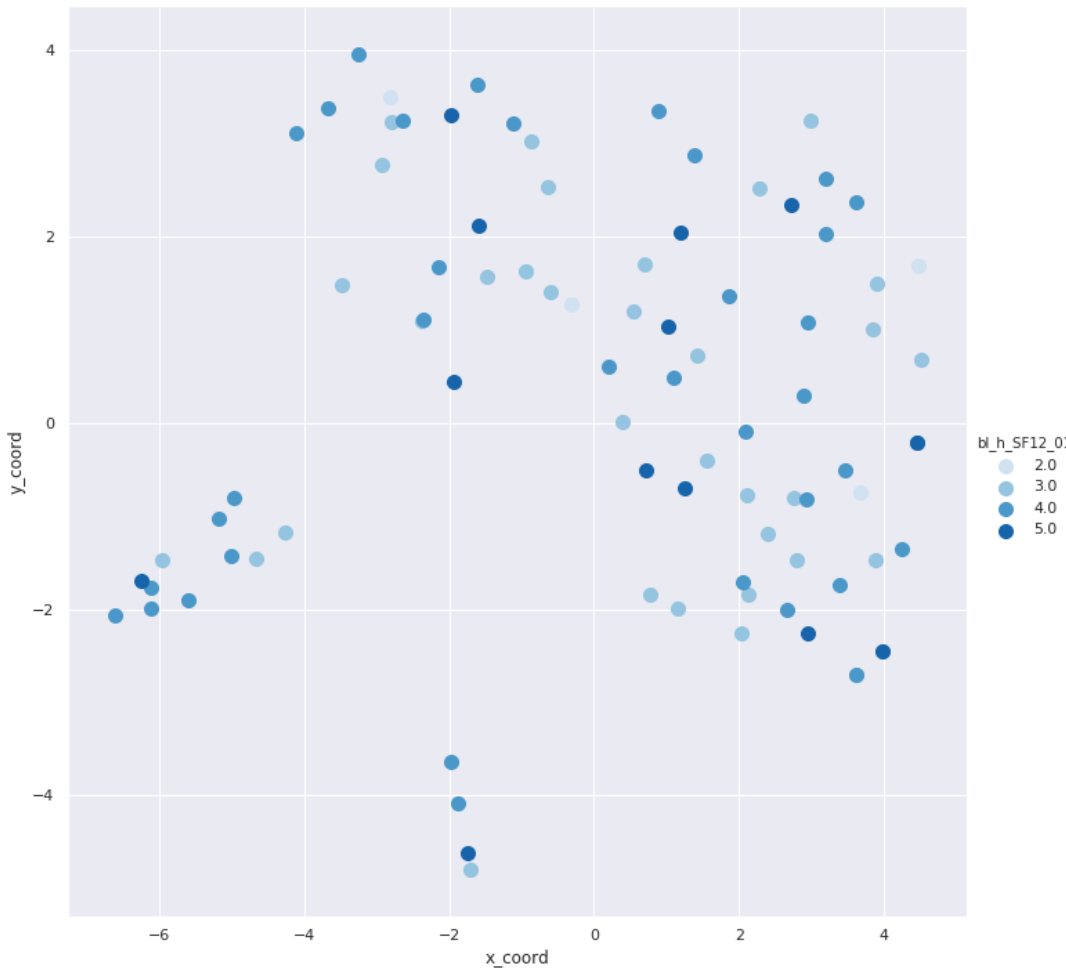

Figure 40 Averaged vectors per participant with the hue based on SF-12_01.

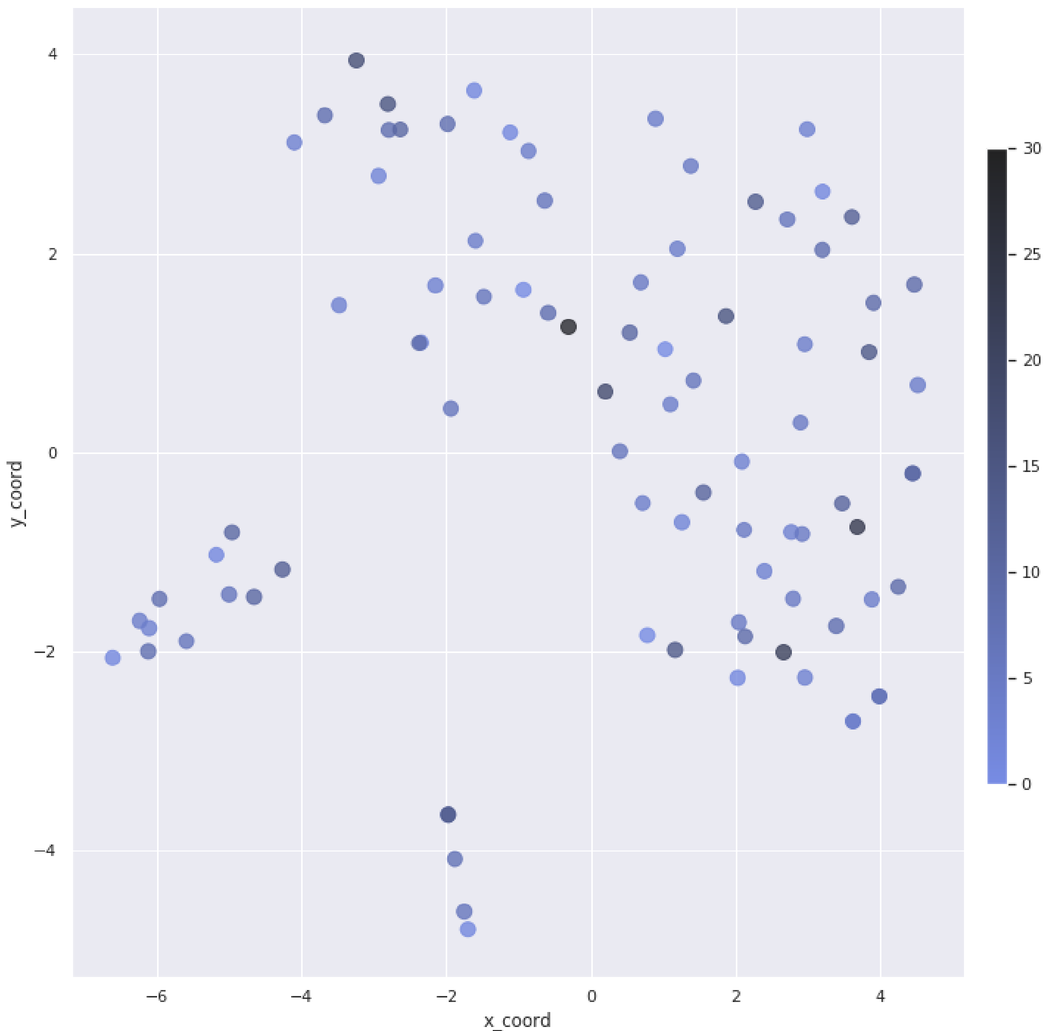

Figure 41 Averaged vectors per participant with the hue based on Depressive symptoms score

6.7 The relationship between common sense clusters and health

The common sense cluster (Appendix C) correlations are in line with the literature. 'Exercise' is associated with lower depressive symptoms and 'Green/blue space' with higher physical health and also with decreased depressive symptoms. The variety of different clusters is associated with decreased physical health, whereas the total number does not have a significant relation. Interestingly eating out is associated with increased depressive symptoms. This could be because eating out-cluster includes fast food. However, as we saw in the individual POI-model, 'restaurant' had a negative association on mental health, whereas 'café' had a positive. It could be that they cancel each other out. Health cluster includes POI types such as 'hospital', 'doctor', and 'pharmacy'. This cluster has an insignificant association with depressive symptoms. However, it was significant before controlling for age, which indicates that these visitations are more related to age rather than depressive symptoms. Many of the clusters were not included in the model after the forward and backward selection. The removed clusters are: ‘Education’, ‘Residential/building’, ‘Retail’, ‘Tourism’, and ‘Public transport’.

	Dependent variable:			
	bl_sf12_mental	bl_sf12_physical	bl_h_SF12_01	blADS
	(1)	(2)	(3)	(4)
Eating out (cube)	-2.072	1.502		1.496*
	(1.452)	(1.130)		(0.906)
Exercise				-0.931**
				(0.419)
Green/blue space (cube)		1.803**		-1.668**
		(0.870)		(0.768)
Health (cube)				2.052
				(1.656)
Miscellaneous (cube)	2.907***			
	(0.788)			
uniquePOIs		-0.896**		
		(0.422)		
total			0.024	
			(0.016)	
age		-0.263**		0.198***
		(0.127)		(0.073)
blsex		1.633		-1.530
		(1.161)		(1.121)
Constant	56.191***	67.153***	3.456***	-7.022
	(0.868)	(9.392)	(0.162)	(5.557)
Observations	88	88	88	88
R2	0.100	0.166	0.023	0.158
Adjusted R2	0.079	0.115	0.012	0.096
Residual Std. Error	5.306 (df = 85)	5.245 (df = 82)	0.798 (df = 86)	4.954 (df = 81)
F Statistic	4.713** (df = 2; 85)	3.270*** (df = 5; 82)	2.064 (df = 1; 86)	2.541** (df = 6; 81)
Note:				*p<0.1; **p<0.05; ***p<0.01

Figure 42 Regression results of the common sense clusters

7 Discussion

Some of the individual POI types are correlated with health, but the investigation of this is highly limited as the data is very skewed. Thus, some clustering method is needed to reduce the skewness. The POI types that were associated with health were ‘fastfood’, ‘forest’, ‘café’, and ‘nature’ with relatively reasonable associations. We could not find proof that the overall frequency of visiting places would be associated with health. This is in line with the literature and that the association between health and place visitations is not a clear linear relationship of ‘visiting more places is better’, which was, however, the core idea of the activity theory. Variety with individual POI types was not significant, but there was an association with the variety of ‘common sense’ clusters with decreased physical health.

Within hierarchical clustering, if there was a relationship within a higher-level cluster and health, this was often similar to at least one of the lower level clusters. Together with the note that the R2 decreases with larger clusters, it could be that this higher-level relationship is mainly influenced only by these few lower-level clusters. Table 5 showed that the skewness reduction can be achieved with relatively small levels of clustering. As we can see the 30-clusters includes most of the POI types, whereas the comparison of individual POI types includes only nine. It is possible that the improvement of the clustering is greatest when combining only a few most similar POI types together to reduce skewness sufficiently.

When using clustering methods, there can be issues in interpreting results. When the clusters are not obvious, this can lead to results where there are significant results, but the reasons for the relationships are very ambiguous, which can result in having noise in the cluster without being aware of this. This can also be an issue with more semantically obvious clusters. For example, when using green spaces as a cluster, it could be that it is a certain type of green space that has a relationship with health, e.g., more rural rather than urban or more social rather than solitudinal space.

For the common sense clusters, which entail only 10 clusters, the R2 values are similar to the ones at 30-clusters, which had the highest R2 values of the hierarchical cluster models. This shows that the hierarchical clustering with Traj2Vec did not improve the semantic modeling of place visitations in relation to health. There are a couple of possible reasons for this: There could be other semantic information that needs to be associated with the places for them to become meaningful. For example, our adapted model of Traj2Vec does not take into account the time of the day or the actual function performed. As we have discussed, there can be many possibilities for semantic enrichment of places. This makes capturing the meaningful part for the analysis challenging. It could also be possible that the semantics of a place attributed by an individual are important. Here we modeled the places based on all of the participants' trajectories, but it is possible that one would need to include a personal element to the model.

The results from Mot2vec showed that the averages of the place vectors across the users did not have a relationship with health. This could be a result of the Traj2Vec modeling and low performance in capturing the important semantics. However, this could also be an indication that it is only a few place types that are associated with older adults' health. This is consistent across the different number of clusters, as only a few place clusters have correlations with the health indicators. It is also in line with the individual POI types and the 'common sense' clusters as many clusters had to be discarded for the best model in forward and backward selections. The few associations with POI types could also be more pronounced in older adults as they are generally less spatially active.

As we predicted, some of the POI types are positively associated with health, whereas others are negatively associated. Across the models, there are some correlations, although many of them are quite weak and inconsistent. Although, many relationships were weak, they seem to be logical, for example, 'café', 'nature, and 'forest' have a positive association with mental health, whereas fast food has a negative association. Visitation frequencies to hospital and social facilities were found to be negatively associated with physical health in some clusters, but they did not hold the significance in a smaller cluster, which would indicate that the significance might be resulting from the other POI types. Green spaces seemed to be consistently associated with increased mental health overall, which is shown in the individual POI, small hierarchical clusters, and the ‘common sense’ clusters. The green space elements were more often associated with increased mental health than physical health and were prevalent even when controlling for age. Taken into account that participation in the study required a certain level of mobility, it is unlikely that increased visitation of green spaces would be associated with higher mobility. There is an abundance of literature on green spaces and health, whereas for other places there are only few studies. This could also be an indication most other places do not have such clear correlations.

It is very likely that there are other semantics that need to be explored to discover correlations between place types and health. There seems to be correlations between certain types of places and health, but the question still arises of which places and what semantics are important to find this correlation. It could be that in place modeling in relation to health, the personal attitudes matter more than the shared semantics of a place, or the sense of place more than the affordance.

8 Limitations

8.1 GPS sensor data quality

The usage of GPS sensors poses multiple limitations. The sensors have problems with accuracy and missing observations. In the data, there were big observation gaps due to the missing GPS waypoints that caused some participants to have less valid daily data, which is a common problem in similar studies (Chaix, 2018). We combated this by using six random days of sufficient data from each person for the statistical part. Increasing the number of days to 12 decreased the participant numbers to 53 and did not increase the POI visitations enough to justify the reduction in participant variability. The statistics of the included vs. excluded people show there is no bias between the two groups. However, the exclusion of the participants might have reduced the variability in the sample. This is also likely to have caused higher skewness in the place visitation rates than would have been in a more long-term sample. The Traj2Vec modeling could have also been be affected by the limited number of days that were included to the model. For the Traj2Vec modeling we used all the days that had sufficient data which increased the available data somewhat in comparison to the statistical sample.

GPS devices’ have some uncertainties in the measuremenst, which are usually more prevalent near or inside buildings, which can cause a bias in the data with higher uncertainties during a place visitation in comparison to transporting or waiting at a bus stop. Lacking of a ground truth for where the participants actually visited during the experiments, we tried to infer the visitations by using relative counts to enrich the semantics of the stop episodes. We included the polygon POI if most of the stop points were inside it, because if there were more stop points outside a building than inside, it is more likely that the actual stop episode was also outside the building. However, a person stopping by a building could be linked to a building in a case when the GPS measurements are biased.

8.2 Incompleteness of OSM POIs

OSM POI data has gaps in terms of attribute annotation, geometry, and location. For example, the tag ‘building’ often does not entail other information, even if there are one or multiple functions. The data is also sparser in the rural regions, which leads to lower granularities in the rural regions. Because OSM data is user-generated, the data can be unsystematic or missing entirely, leading to the selection of the wrong POIs during the stop episode enrichment.

8.3 Health indicators

The health indicators are slightly skewed, which is most likely due to the sampling. The nature of the experiment required the participants to be relatively healthy with adequate mobility levels. When combined with the sample size of 159, this results in slightly skewed sample towards a healthier sample than would be otherwise. It is likely, that older adults with better health would have higher mobility, and hence, the accessibility to places would not be accounted for in this study. Thus, the results must be understood in the light that the mobility of the participants was relatively good. Thus the regression analysis would be rather showing the differences within healthy older adults. It is also possible that the skewed indicators affected the regression results.

8.4 The propagation of data quality to modeling

The noise in the raw trajectory data might have had an impact on the final models. The trajectory data were decreased with the selection of the valid days for the Traj2Vec modeling. As there was insufficient data input, the model might be undertrained. As our model assumes that the semantics are the same across the participants, we used all the valid days to maximize the sampling size. However, the reduced sample of valid days might have skewed the modeling of the Traj2Vec to certain participants’ trajectories, i.e., some of the semantics might have been mainly modeled only by few participants’ the trajectories.

The gaps in the GPS data may have propagated to trajectory segmentation. Although the gaps were minimized and only valid days were taken for the analysis, some gaps in the data remain, and these may have resulted in missing stop episodes in the trajectories. This might have then carried to place embedding and the POI counts.

The gaps in the OSM data might have have caused some issues in the semantic enrichment, which would have carried to the place modeling. If a POI had a wrong semantic label, due to missing or wrong data, this would have affected the Traj2Vec modeling. Generally, the OSM data is more accurate, has higher spatial coverage, and richer semantic information in the cities. Switzerland also has many open sources for the infrastructure data, which is normally incorporated in the OSM data as well. However, especially in rural areas, the data can be sparser, which can slightly distort the trajectories that are further from cities.

8.5 Limitations due to movement modeling

There is variability in the data in the types of movements and stops which makes the trajectory segmentation challenging. For example, walking in a park or walking to a swimming pool can be very similar in their physical attributes and thus difficult to separate into stops and moves. As we discussed in the literature review, many researchers focus on one type of movement and the separation of moves based on known movement types, such as driving or public transportation. There is no best practice for the segmentation of the trajectory with this type of data and the results can vary based on the nature of the trajectories. We used POSMIT with automatic variable detection to detect each person's typical stops aiming to ensure that we do not exclude frequent places that are commonly visited by a person.

The stop episodes are formed from waypoint clusters with a variety of densities in space. For example, a stop episode at a park might have a much lower density than at a corner store. This has implications on the semantic enrichment. Many of the existing semantic enrichment studies have tackled within urban POI types and note that the effectiveness of different methods varies depending on the density of the POI types. The semantic enrichment of this study was done relatively 'cautiously' by setting the count limits high and distance limits fairly low. Because a high number of the observations was in relatively dense areas, a relaxation in the rules would have led to more misclassified POIs. The drawback is that this often leads the stops being linked with land use polygons and thus to a lower granularity. Future research could investigate different POI enrichment methods in such diverse movement data.

9 Conclusions and future research

We have investigated the relationship between health and place visitations and found that there are some places that have a positive association, e.g., green spaces and places related to exercises, while others, e.g., fast food restaurants, have a negative association with health, which is in line with our hypothesis. However, as predicted, it is not a simple linear correlation but varies between different places and is especially pronounced in some types in comparison to others. Green spaces seemed to be most consistently associated with increased health indicators. Other examples of place types related to health are exercise or health care related.

The individual place type -based model had interesting findings: 'forest', 'nature', and 'café' were associated with increased mental health and 'fast food' with decreased physical health. However, the results were highly limited due to the high skewness of the place visitations. The influence of the individual POI types was carried to the embedding-based clustering models with similar associations. These findings were relatively similar to the 'common sense' clusters as well. In contrary to an initial prediction, the variability and the total number of visited places did not have an association with health, which could be an indication that only few place types have a relationship with health. As could be seen from the 'common sense' multivariate regressions, these commonly associated semantics can encapsulate some relationships between health and place visitations.

Between the different health indicators, there are some differences. Both mental and physical health seem to be positively associated with green space visitations, based on different regression results. Depressive symptoms are also negatively associated with exercise and positively with eating out. Indicator

SF-12_01, which is the overall self-assessed health, seems to have the least associations with the place visitations. This could imply that the associations between place visitations and health are not based on how the person perceives their mobility and health.

Methodology-wise, we aimed to model the semantics of places using Traj2Vec, which however did not increase the performance of our models. The ‘common sense’ clusters of places performed better in terms of their explanatory power, interpretability, and significance. However, the comparisons of the individual POI-model and the cluster models show the importance of clustering as means of reducing skewness in the place visitations.

Thus, we call for more research in extracting the meaningful semantics of the places to discover the relationship to health. The modeling of the places could also be done through another method than Traj2Vec, such as place2vec or habit2vec. Place2vec would require some additional data, such as the visitations of all places, which are not available for OSM data. Our models show that there are some relationships with the visitation of certain places. Therefore it could be, that the Traj2Vec was not able to capture this connection, but some other models might.

Other information could be incorporated into the data, possibly to improve the models. Some place models use grids to combine similar places. For this research, it would have been challenging as there is a great variety in terms of the data density over the regions. One approach could be to take the data only from the Zurich region and divide this into grids. However, this would reduce the data further and could leave out some important features, such as recreational visits to other parts of Switzerland. Another approach would be to incorporate the time span of the ‘sentences’, i.e., including the temporal window into the segmentation. In this research, this was not conducted as the stop episodes were detected to get the correct frequency rather than the time span, which might have reduced the temporal size of some stops. Also, as the data was relatively sparse, this would have required a more consistent temporal match between the subjects.

Some of the semantics could not be derived from the data. For example, social interactions have been found to be an important factor in health research. However, it is not very straightforward to derive from the data as places most often do not entail sociality per se. In future research, the combination of places and their sociality could be further explored.

This study is concerned with modeling the affordances of places, but it could be that the health is more related to places with the definition of ‘sense of place’ where the individual attitudes and understandings of the place formulate the relationship rather than the function. Within the theory of affordance, the function that a place provides can differ between groups and individuals, as discussed in the literature. However, Traj2Vec and Mot2vec incorporate semantics as a global variable. Bin et al. (2020) found that Traj2Vec performs relatively well when using the group as an indicator but performs even better when using a personalised model. Here the Traj2Vec modeling was done by aggregating a person’s all visitations as sentences, but if there was a larger data, perhaps the skewness could be corrected by aggregating days as sentences, which might change the clustering.

The models might be improved with additional data. The mobility behavior could have been affected by other environmental factors besides what was accounted for in this study, such as accessibility, socioeconomic status, etc. The POI detection models have also been called for using habit data to further analyze the relationship with different other parameters such as weather, time of the day, weekday, season, etc. (Krueger et al., 2015). For example, Gong et al. (2018) found that seasonality is important when training a machine learning algorithm to detect stop purposes.

As discussed above, there are many ways in which the place visitation models could be improved. The models aim to capture the meaningful semantics of the places, which can be highly challenging as there can be an infinite number of semantics embedded in a place, but only some are meaningful. The meaningfulness of the semantics also depends on the analysis, and the specific semantics might be meaningful in one particular analysis but not in another. We found that there are correlations between health in older adults and visitations of some place types. However, this relationship is complex and requires further research.

10 Acknowledgments

I would like to thank my supervisor Cheng Fu, for explaining difficult concepts in understandable terms and mentoring me all the way, which has been a great support; Robert Weibel, for important advice and help; Christina Röcke, for the collaboration and enlightening psychology faculty perspective; Eun-Kyeong Kim, for the guidance about the data structure; Elena Ebert, for her advice and contributions. Tobias Hoffmann, for helping with technical issues; My partner and my family, for supporting me on my journey. Along with many others who assisted me throughout the way.

11 Literature

Alessandretti, L., Lehmann, S., Baronchelli, A., 2018. Understanding the interplay between social and spatial behaviour. EPJ Data Sci. 7, 36. https://doi.org/10.1140/epjds/s13688-018-0164-6

Alvares, L.O., Bogorny, V., Kuijpers, B., de Macedo, J.A.F., Moelans, B., Vaisman, A., 2007. A model for enriching trajectories with semantic geographical information, in: Proceedings of the 15th Annual ACM International Symposium on Advances in Geographic Information Systems, GIS '07. Association for Computing Machinery, Seattle, Washington, pp. 1–8. https://doi.org/10.1145/1341012.1341041

Alvarez-Garcia, J.A., Ortega, J.A., Gonzalez-Abril, L., Velasco, F., 2010. Trip destination prediction based on past GPS log using a Hidden Markov Model. Expert Syst. Appl. 37, 8166–8171. https://doi.org/10.1016/j.eswa.2010.05.070

Ankerst, M., Breunig, M.M., Kriegel, H.-P., Sander, J., 1999. OPTICS: Ordering Points to Identify the Clustering Structure, in: Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data, SIGMOD '99. ACM, New York, NY, USA, pp. 49–60. https://doi.org/10.1145/304182.304187

Baldassare, M., Rosenfield, S., Rook, K., 1984. The Types of Social Relations Predicting Elderly Well-Being. Res. Aging 6, 549–559. https://doi.org/10.1177/0164027584006004006

Barrett, B., Nitze, I., Green, S., Cawkwell, F., 2014. Assessment of multi-temporal, multi-sensor radar and ancillary spatial data for grasslands monitoring in Ireland using machine learning approaches. Remote Sens. Environ. 152, 109–124. https://doi.org/10.1016/j.rse.2014.05.018

Bermingham, L., Lee, I., 2019. Mining place-matching patterns from spatio-temporal trajectories using complex real-world places. Expert Syst. Appl. 122, 334–350. https://doi.org/10.1016/j.eswa.2019.01.027

Bermingham, L., Lee, I., 2018. A probabilistic stop and move classifier for noisy GPS trajectories. Data Min. Knowl. Discov. 32, 1634–1662. https://doi.org/10.1007/s10618-018-0568-8

Biljecki, F., Ledoux, H., Oosterom, P. van, 2013. Transportation mode-based segmentation and classification of movement trajectories. Int. J. Geogr. Inf. Sci. 27, 385–407. https://doi.org/10.1080/13658816.2012.692791

Bin, C., Gu, T., Jia, Z., Zhu, G., Xiao, C., 2020. A neural multi-context modeling framework for personalized attraction recommendation. Multimed. Tools Appl. 79, 14951–14979. https://doi.org/10.1007/s11042-019-08554-5

Birant, D., Kut, A., 2007. ST-DBSCAN: An algorithm for clustering spatial–temporal data. Data Knowl. Eng., Intelligent Data Mining 60, 208–221. https://doi.org/10.1016/j.datak.2006.01.013

Bohte, W., Maat, K., 2009. Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. Transp. Res. Part C Emerg. Technol. 17, 285–297. https://doi.org/10.1016/j.trc.2008.11.004

Bond, J., Peace, S.M., Dittmann-Kohli, F., Westerhof, G., 2007. Ageing in Society. SAGE.

Brakatsoulas, S., Pfoser, D., Tryfona, N., 2004. Modeling, storing and mining moving object databases, in: Proceedings. International Database Engineering and Applications Symposium, 2004. IDEAS ’04. Presented at the Proceedings. International Database Engineering and Applications Symposium, 2004. IDEAS ’04., pp. 68–77. https://doi.org/10.1109/IDEAS.2004.1319779

Buchin, M., Driemel, A., Kreveld, M. van, Sacristan, V., 2011. Segmenting trajectories: A framework and algorithms using spatiotemporal criteria. J. Spat. Inf. Sci. 2011, 33–63. https://doi.org/10.5311/JOSIS.2011.3.66

Cao, H., Xu, F., Sankaranarayanan, J., Li, Y., Samet, H., 2020. Habit2vec: Trajectory Semantic Embedding for Living Pattern Recognition in Population. IEEE Trans. Mob. Comput. 19, 1096–1108. https://doi.org/10.1109/TMC.2019.2902403

Celli, F., Spinsanti, L., Renso, C., 2010. Where you stop is who you are: Understanding people’s activities by places visited.

Chaix, B., 2018. Mobile Sensing in Environmental Health and Neighborhood Research. Annu. Rev. Public Health 39, 367–384. https://doi.org/10.1146/annurev-publhealth-040617-013731

Chen, W., Pourghasemi, H.R., Kornejady, A., Zhang, N., 2017. Landslide spatial modeling: Introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. Geoderma 305, 314–327. https://doi.org/10.1016/j.geoderma.2017.06.020

Cox, D.T.C., Shanahan, D.F., Hudson, H.L., Plummer, K.E., Siriwardena, G.M., Fuller, R.A., Anderson, K., Hancock, S., Gaston, K.J., 2017. Doses of Neighborhood Nature: The Benefits for Mental Health of Living with Nature. BioScience 67, 147–155. https://doi.org/10.1093/biosci/biw173

Crivellari, A., Beinat, E., 2019. From Motion Activity to Geo-Embeddings: Generating and Exploring Vector Representations of Locations, Traces and Visitors through Large-Scale Mobility Data. ISPRS Int. J. Geo-Inf. 8, 134. https://doi.org/10.3390/ijgi8030134

Dallimer, M., Irvine, K.N., Skinner, A.M.J., Davies, Z.G., Rouquette, J.R., Maltby, L.L., Warren, P.H., Armsworth, P.R., Gaston, K.J., 2012. Biodiversity and the Feel-Good Factor: Understanding

Associations between Self-Reported Human Well-being and Species Richness. BioScience 62, 47–55. https://doi.org/10.1525/bio.2012.62.1.9

Das, R.D., Winter, S., 2016. Automated Urban Travel Interpretation: A Bottom-up Approach for Trajectory Segmentation. Sensors 16, 1962. https://doi.org/10.3390/s16111962

de Graaff, V., de By, R.A., van Keulen, M., 2016. Automated semantic trajectory annotation with indoor point-of-interest visits in urban areas, in: Proceedings of the 31st Annual ACM Symposium on Applied Computing, SAC '16. Association for Computing Machinery, Pisa, Italy, pp. 552–559. https://doi.org/10.1145/2851613.2851709

de Graaff, V., de By, R.A., van Keulen, M., Flokstra, J., 2013. Point of interest to region of interest conversion, in: Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, SIGSPATIAL'13. Association for Computing Machinery, Orlando, Florida, pp. 388–391. https://doi.org/10.1145/2525314.2525442

Deng, D.-P., Chuang, T.-R., Lemmens, R., 2009. Conceptualization of Place via Spatial Clustering and Co-occurrence Analysis, in: Proceedings of the 2009 International Workshop on Location Based Social Networks, LBSN '09. ACM, New York, NY, USA, pp. 49–56. https://doi.org/10.1145/1629890.1629902

Enam, A., Konduri, K.C., Eluru, N., Ravulaparthy, S., 2018. Relationship between well-being and daily time use of elderly: evidence from the disabilities and use of time survey. Transportation 45, 1783–1810.

Farquhar, M., 1995. Elderly people's definitions of quality of life. Soc. Sci. Med., Quality of Life in Social Science and Medicine 41, 1439–1446. https://doi.org/10.1016/0277-9536(95)00117-P

Feng, T., Timmermans, H.J.P., 2016. Comparison of advanced imputation algorithms for detection of transportation mode and activity episode using GPS data. Transp. Plan. Technol. 39, 180–194. https://doi.org/10.1080/03081060.2015.1127540

Fileto, R., Krüger, M., Pelekis, N., Theodoridis, Y., Renso, C., 2013. Baquara: A Holistic Ontological Framework for Movement Analysis Using Linked Data, in: Ng, W., Storey, V.C., Trujillo, J.C. (Eds.), Conceptual Modeling, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 342–355. https://doi.org/10.1007/978-3-642-41924-9_28

Fillekes, M.P., Giannouli, E., Kim, E.-K., Zijlstra, W., Weibel, R., 2019. Towards a comprehensive set of GPS-based indicators reflecting the multidimensional nature of daily mobility for applications in health and aging research. Int. J. Health Geogr. 18. https://doi.org/10.1186/s12942-019-0181-0

Fonseca, F., Egenhofer, M., Davis, C., Câmara, G., 2002. Semantic Granularity in Ontology-Driven Geographic Information Systems. Ann. Math. Artif. Intell. 36, 121–151. https://doi.org/10.1023/A:1015808104769

Garg, N., Ramadurai, G., Ranu, S., 2018. Mining bus stops from raw GPS data of bus trajectories, in: 2018 10th International Conference on Communication Systems Networks (COMSNETS). Presented at the 2018 10th International Conference on Communication Systems Networks (COMSNETS), pp. 583–588. https://doi.org/10.1109/COMSNETS.2018.8328278

Geofabrik Download Server [WWW Document], 2020. URL https://download.geofabrik.de/europe/switzerland.html (accessed 7.29.20).

Gibson, J.J., 1979. The ecological approach to visual perception. Boston, MA, US. Houghton, Mifflin and Company.

Gong, L., Kanamori, R., Yamamoto, T., 2018a. Data selection in machine learning for identifying trip purposes and travel modes from longitudinal GPS data collection lasting for seasons. Travel Behav. Soc. 11, 131–140. https://doi.org/10.1016/j.tbs.2017.03.004

Gong, L., Sato, H., Yamamoto, T., Miwa, T., Morikawa, T., 2015. Identification of activity stop locations in GPS trajectories by density-based clustering method combined with support vector machines. J. Mod. Transp. 23, 202–213. https://doi.org/10.1007/s40534-015-0079-x

Gong, L., Yamamoto, T., Morikawa, T., 2018b. Identification of activity stop locations in GPS trajectories by DBSCAN-TE method combined with support vector machines. Transp. Res. Procedia 32, 146–154. https://doi.org/10.1016/j.trpro.2018.10.028

Greider, T., Garkovich, L., 1994. Landscapes: The Social Construction of Nature and the Environment. Rural Sociol. 59, 1–24. https://doi.org/10.1111/j.1549-0831.1994.tb00519.x

Gu, Q., Sacharidis, D., Mathioudakis, M., Wang, G., 2017. Inferring Venue Visits from GPS Trajectories, in: Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, SIGSPATIAL '17. Association for Computing Machinery, Redondo Beach, CA, USA, pp. 1–4. https://doi.org/10.1145/3139958.3140034

Habib, K.M.N., Hui, V., 2017. An activity-based approach of investigating travel behaviour of older people. Transportation 44, 555–573. https://doi.org/10.1007/s11116-015-9667-1

Handler, S., 2014. An Alternative Age-friendly\copyright Handbook: Featuring New & Emerging Age-inclusive Initiatives with Additional Reflections (& Provocations) on Ageing in the Contemporary City. University of Manchester Library.

Harris, Z.S., 1954. Distributional Structure. *WORD* 10, 146–162. https://doi.org/10.1080/00437956.1954.11659520

Hornsby, K., Egenhofer, M.J., 2002. Modeling Moving Objects over Multiple Granularities 18.

Huang, C.-M., Ying, J.J.-C., Tseng, V.S., 2012. Mining Users' Behaviors and Environments for Semantic Place Prediction [WWW Document]. undefined. URL /paper/Mining-Users%27-Behaviors-and-Environments-for-Place-Huang-Ying/bf643741d5a5e8d83202ca92c931c59bc919c7cd (accessed 7.22.20).

Hummon, D.M., 1992. Community Attachment, in: Altman, I., Low, S.M. (Eds.), Place Attachment, Human Behavior and Environment. Springer US, Boston, MA, pp. 253–278. https://doi.org/10.1007/978-1-4684-8753-4_12

Hwang, S., Evans, C., Hanke, T., 2017. Detecting Stop Episodes from GPS Trajectories with Gaps, in: Thakuriah, P. (Vonu), Tilahun, N., Zellner, M. (Eds.), Seeing Cities Through Big Data: Research, Methods and Applications in Urban Informatics, Springer Geography. Springer International Publishing, Cham, pp. 427–439. https://doi.org/10.1007/978-3-319-40902-3_23

Hwang, S., VanDeMark, C., Dhatt, N., Yalla, S.V., Crews, R.T., 2018. Segmenting human trajectory data by movement states while addressing signal loss and signal noise. Int. J. Geogr. Inf. Sci. 32, 1391–1412. https://doi.org/10.1080/13658816.2018.1423685

Inglese, P., S. McKenzie, J., Mroz, A., Kinross, J., Veselkov, K., Holmes, E., Takats, Z., K. Nicholson, J., C. Glen, R., 2017. Deep learning and 3D-DESI imaging reveal the hidden metabolic heterogeneity of cancer. Chem. Sci. 8, 3500–3511. https://doi.org/10.1039/C6SC03738K

Isaacson, M., Wahl, H.-W., Shoval, N., Oswald, F., Auslander, G., 2017. The Relationship Between Spatial Activity and Wellbeing-Related Data Among Healthy Older Adults: An Exploratory Geographic and Psychological Analysis, in: Cross-Cultural and Cross-Disciplinary Perspectives in Social Gerontology. pp. 203–219. https://doi.org/10.1007/978-981-10-1654-7_11

Krueger, R., Thom, D., Ertl, T., 2015. Semantic Enrichment of Movement Behavior with Foursquare–A Visual Analytics Approach. IEEE Trans. Vis. Comput. Graph. 21, 903–915. https://doi.org/10.1109/TVCG.2014.2371856

Kwan, M.P., Lee, J., 2003. Geovisualization of human activity patterns using 3D GIS: a time-geographic approach, in: Spatially Integrated Social Science: Examples in Best Practice. Oxford University Press.

Kyle, G., Chick, G., 2007. The Social Construction of a Sense of Place. Leis. Sci. 29, 209–225. https://doi.org/10.1080/01490400701257922

Lehmann, A.L., Alvares, L.O., Bogorny, V., 2019. SMSM: a similarity measure for trajectory stops and moves. Int. J. Geogr. Inf. Sci. 33, 1847–1872. https://doi.org/10.1080/13658816.2019.1605074

Lemon, B.W., Bengtson, V.L., Peterson, J.A., 1972. An Exploration of the Activity Theory of Aging: Activity Types and Life Satisfaction Among In-movers to a Retirement Community. J. Gerontol. 27, 511–523. https://doi.org/10.1093/geronj/27.4.511

Lin, B.B., Fuller, R.A., Bush, R., Gaston, K.J., Shanahan, D.F., 2014. Opportunity or Orientation? Who Uses Urban Parks and Why. PLOS ONE 9, e87422. https://doi.org/10.1371/journal.pone.0087422

Lin, Q., Zhang, D., Huang, X., Ni, H., Zhou, X., 2012. Detecting wandering behavior based on GPS traces for elders with dementia, in: 2012 12th International Conference on Control Automation Robotics Vision (ICARCV). Presented at the 2012 12th International Conference on Control Automation Robotics Vision (ICARCV), pp. 672–677. https://doi.org/10.1109/ICARCV.2012.6485238

Litwin, H., Shiovitz-Ezra, S., 2006. The association between activity and wellbeing in later life: what really matters? Ageing Soc. 26, 225–242. https://doi.org/10.1017/S0144686X05004538

Longino, C.F., Kart, C.S., 1982. Explicating Activity Theory: A Formal Replication. J. Gerontol. 37, 713–722. https://doi.org/10.1093/geronj/37.6.713

Luo, T., Zheng, X., Xu, G., Fu, K., Ren, W., 2017. An Improved DBSCAN Algorithm to Detect Stops in Individual Trajectories. ISPRS Int. J. Geo-Inf. 6, 63. https://doi.org/10.3390/ijgi6030063

Lv, M., Chen, L., Xu, Z., Li, Y., Chen, G., 2016. The discovery of personally semantic places based on trajectory data mining. Neurocomputing 173, 1142–1153. https://doi.org/10.1016/j.neucom.2015.08.071

Map Features – OpenStreetMap Wiki [WWW Document], n.d. URL https://wiki.openstreetmap.org/wiki/Map_Features (accessed 8.5.20).

Mark, D.M., Freksa, C., Hirtle, S.C., Lloyd, R., Tversky, B., 1999. Cognitive models of geographical space. Int. J. Geogr. Inf. Sci. 13, 747–774. https://doi.org/10.1080/136588199241003

Marquet, O., Miralles-Guasch, C., 2015. Neighbourhood vitality and physical activity among the elderly: The role of walkable environments on active ageing in Barcelona, Spain. Soc. Sci. Med. 135, 24–30. https://doi.org/10.1016/j.socscimed.2015.04.016

McKercher, B., Shoval, N., Ng, E., Birenboim, A., 2012. First and Repeat Visitor Behaviour: GPS Tracking and GIS Analysis in Hong Kong. Tour. Geogr. 14, 147–161. https://doi.org/10.1080/14616688.2011.598542

Meng, C., Cui, Y., He, Q., Su, L., Gao, J., 2019. Towards the Inference of Travel Purpose with Heterogeneous Urban Data. IEEE Trans. Big Data 1–1. https://doi.org/10.1109/TBDATA.2019.2921823

Meng, C., Cui, Y., He, Q., Su, L., Gao, J., 2017. Travel purpose inference with GPS trajectories, POIs, and geo-tagged social media data, in: 2017 IEEE International Conference on Big Data (Big Data). Presented at the 2017 IEEE International Conference on Big Data (Big Data), pp. 1319–1324. https://doi.org/10.1109/BigData.2017.8258062

Mikolov, T., Chen, K., Corrado, G., Dean, J., 2013. Efficient Estimation of Word Representations in Vector Space. ArXiv13013781 Cs.

Mollenkopf, H., Hieber, A., Wahl, H.-W., 2011. Continuity and change in older adults' perceptions of out-of-home mobility over ten years: a qualitative–quantitative approach. Ageing Soc. 31, 782–802. https://doi.org/10.1017/S0144686X10000644

Mollenkopf, H., Marcellini, F., Ruoppila, I., Széman, Z., Tacken, M., Wahl, H.-W., 2004. Social and behavioural science perspectives on out-of-home mobility in later life: findings from the European project MOBILATE. Eur. J. Ageing 1, 45–53. https://doi.org/10.1007/s10433-004-0004-3

Moreno, F.J., Pineda, A.F., Fileto, R., Bogorny, V., 2014. SMoT+: Extending the SMoT Algorithm for Discovering Stops in Nested Sites. Comput. Inform. 33, 327-342–342.

Nimmo, R.D., Callender, G., Lewin, P.L., 2017. Methods for wavelet-based autonomous discrimination of multiple partial discharge sources. IEEE Trans. Dielectr. Electr. Insul. 24, 1131–1140. https://doi.org/10.1109/TDEI.2017.006157

Nutsford, D., Pearson, A.L., Kingham, S., Reitsma, F., 2016. Residential exposure to visible blue space (but not green space) associated with lower psychological distress in a capital city. Health Place 39, 70–78. https://doi.org/10.1016/j.healthplace.2016.03.002

Ottoni, C.A., Sims-Gould, J., Winters, M., Heijnen, M., McKay, H.A., 2016. "Benches become like porches": Built and social environment influences on older adults' experiences of mobility and well-being. Soc. Sci. Med. 169, 33–41. https://doi.org/10.1016/j.socscimed.2016.08.044

Ottosson, J., Grahn, P., 2005. A Comparison of Leisure Time Spent in a Garden with Leisure Time Spent Indoors: On Measures of Restoration in Residents in Geriatric Care. Landsc. Res. 30, 23–55. https://doi.org/10.1080/0142639042000324758

Palma, A.T., Bogorny, V., Kuijpers, B., Alvares, L.O., 2008. A clustering-based approach for discovering interesting places in trajectories, in: Proceedings of the 2008 ACM Symposium on Applied Computing, SAC '08. Association for Computing Machinery, Fortaleza, Ceara, Brazil, pp. 863–868. https://doi.org/10.1145/1363686.1363886

Pandove, D., Goel, S., Rani, R., 2018. Systematic Review of Clustering High-Dimensional and Large Datasets. ACM Trans. Knowl. Discov. Data 12, 16:1–16:68. https://doi.org/10.1145/3132088

Parent, C., Spaccapietra, S., Renso, C., Andrienko, G., Andrienko, N., Bogorny, V., Damiani, M.L., Gkoulalas-Divanis, A., Macedo, J., Pelekis, N., Theodoridis, Y., Yan, Z., 2013. Semantic trajectories modeling and analysis. ACM Comput. Surv. 45, 42:1–42:32. https://doi.org/10.1145/2501654.2501656

Ruta, M., Scioscia, F., Ieva, S., Loseto, G., Di Sciascio, E., 2012. Semantic Annotation of OpenStreetMap Points of Interest for Mobile Discovery and Navigation. Presented at the Proceedings - 2012 IEEE 1st International Conference on Mobile Services, MS 2012, pp. 33–39. https://doi.org/10.1109/MobServ.2012.17

Schratz, P., Muenchow, J., Iturritxa, E., Richter, J., Brenning, A., 2019. Performance evaluation and hyperparameter tuning of statistical and machine-learning models using spatial data. Ecol. Model. 406, 109–120. https://doi.org/10.1016/j.ecolmodel.2019.06.002

Schuessler, N., Axhausen, K.W., 2009. Processing Raw Data from Global Positioning Systems without Additional Information: Transp. Res. Rec. https://doi.org/10.3141/2105-04

Schwanen, T., Ziegler, F., 2011. Wellbeing, independence and mobility: an introduction. Ageing Soc. 31.
Shi, H., Li, Y., Cao, H., Zhou, X., Zhang, C., Kostakos, V., 2019a. Semantics-Aware Hidden Markov Model for Human Mobility. IEEE Trans. Knowl. Data Eng. 1–1. https://doi.org/10.1109/TKDE.2019.2937296
Shi, H., Zhang, C., Yao, Q., Li, Y., Sun, F., Jin, D., 2019b. State-Sharing Sparse Hidden Markov Models for Personalized Sequences, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '19. Association for Computing Machinery, Anchorage, AK, USA, pp. 1549–1559. https://doi.org/10.1145/3292500.3330828
Shi, J., Mamoulis, N., Wu, D., Cheung, D.W., 2014. Density-based Place Clustering in Geo-social Networks, in: Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, SIGMOD '14. ACM, New York, NY, USA, pp. 99–110. https://doi.org/10.1145/2588555.2610497
Shoval, N., Auslander, G., Cohen-Shalom, K., Isaacson, M., Landau, R., Heinik, J., 2010. What can we learn about the mobility of the elderly in the GPS era? J. Transp. Geogr., Special issue: The mobility of older people 18, 603–612. https://doi.org/10.1016/j.jtrangeo.2010.03.012
Shoval, N., Isaacson, M., 2007. Sequence alignment as a method for human activity analysis in space and time. Ann. Assoc. Am. Geogr. 97, 282–297. https://doi.org/10.1111/j.1467-8306.2007.00536.x
Stedman, R.C., 2003. Is It Really Just a Social Construction?: The Contribution of the Physical Environment to Sense of Place. Soc. Nat. Resour. 16, 671–685. https://doi.org/10.1080/08941920309189
Steiger, E., Resch, B., Zipf, A., 2016. Exploration of spatiotemporal and semantic clusters of Twitter data using unsupervised neural networks. Int. J. Geogr. Inf. Sci. 30, 1694–1716. https://doi.org/10.1080/13658816.2015.1099658
Stokowski, P.A., 2002. Languages of Place and Discourses of Power: Constructing New Senses of Place. J. Leis. Res. 34, 368–382. https://doi.org/10.1080/00222216.2002.11949977
Stopher, P., FitzGerald, C., Zhang, J., 2008. Search for a global positioning system device to measure person travel. Transp. Res. Part C Emerg. Technol., Emerging Commercial Technologies 16, 350–369. https://doi.org/10.1016/j.trc.2007.10.002
Tobler, W.R., 1970. Movie Simulating Urban Growth in the Detroit Region Author (s) :
Trajcevski, G., 2011. Uncertainty in Spatial Trajectories, in: Zheng, Y., Zhou, X. (Eds.), Computing with Spatial Trajectories. Springer, New York, NY, pp. 63–107. https://doi.org/10.1007/978-1-4614-1629-6_3
Tran, L.H., Nguyen, Q.V.H., Do, N.H., Yan, Z., 2011. Robust and Hierarchical Stop Discovery in Sparse and Diverse Trajectories [WWW Document]. Infoscience. URL https://infoscience.epfl.ch/record/175473 (accessed 7.17.20).
Tuan, Y.-F., 1979. Space and Place: Humanistic Perspective, in: Gale, S., Olsson, G. (Eds.), Philosophy in Geography. Springer Netherlands, Dordrecht, pp. 387–427. https://doi.org/10.1007/978-94-009-9394-5_19
Tuan, Y.-F., 1975. Place: An Experiential Perspective. Geogr. Rev. 65, 151–165. https://doi.org/10.2307/213970

Tung, F., Zelek, J.S., Clausi, D.A., 2011. Goal-based trajectory analysis for unusual behaviour detection in intelligent surveillance. Image Vis. Comput. 29, 230–240. https://doi.org/10.1016/j.imavis.2010.11.003

Turney, P.D., Pantel, P., 2010. From Frequency to Meaning: Vector Space Models of Semantics. J. Artif. Intell. Res. 37, 141–188. https://doi.org/10.1613/jair.2934

van den Berg, M.M., van Poppel, M., van Kamp, I., Ruijsbroek, A., Triguero-Mas, M., Gidlow, C., Nieuwenhuijsen, M.J., Gražulevičiene, R., van Mechelen, W., Kruize, H., Maas, J., 2019. Do Physical Activity, Social Cohesion, and Loneliness Mediate the Association Between Time Spent Visiting Green Space and Mental Health? Environ. Behav. 51, 144–166. https://doi.org/10.1177/0013916517738563

Vulchanova, M., Zee, E. van der, 2013. Motion Encoding in Language and Space. OUP Oxford.

Waga, K., Tabarcea, A., Chen, M., Fränti, P., 2012. Detecting movement type by route segmentation and classification, in: 8th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom). Presented at the 8th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), pp. 508–513. https://doi.org/10.4108/icst.collaboratecom.2012.250450

Wan, N., Lin, G., 2016. Classifying Human Activity Patterns from Smartphone Collected GPS data: A Fuzzy Classification and Aggregation Approach. Trans. GIS 20, 869–886. https://doi.org/10.1111/tgis.12181

Wen, Y., Zhang, Y., Huang, L., Zhou, C., Xiao, C., Zhang, F., Peng, X., Zhan, W., Sui, Z., 2019. Semantic Modelling of Ship Behavior in Harbor Based on Ontology and Dynamic Bayesian Network. ISPRS Int. J. Geo-Inf. 8, 107. https://doi.org/10.3390/ijgi8030107

WHO, 2019. WHO | WHOQOL: Measuring Quality of Life [WWW Document]. Health Stat. Inf. Syst. URL https://www.who.int/healthinfo/survey/whoqol-qualityoflife/en/ (accessed 10.24.19).

WHO, 1948. Constitution [WWW Document]. URL https://www.who.int/about/who-we-are/constitution (accessed 10.24.19).

Wilkie, S., Stavridou, A., 2013. Influence of environmental preference and environment type congruence on judgments of restoration potential. Urban For. Urban Green. 12, 163–170. https://doi.org/10.1016/j.ufug.2013.01.004

Wong, R.C.P., Szeto, W.Y., Yang, L., Li, Y.C., Wong, S.C., 2018. Public transport policy measures for improving elderly mobility. Transp. Policy 63, 73–79. https://doi.org/10.1016/j.tranpol.2017.12.015

Xiang, L., Gao, M., Wu, T., 2016. Extracting Stops from Noisy Trajectories: A Sequence Oriented Clustering Approach. ISPRS Int J Geo-Inf. 5, 29. https://doi.org/10.3390/ijgi5030029

Xiao, H., Wang, W.J., Zhang, X., 2013. Identifying the Stay Point Using GPS Trajectory of Taxis [WWW Document]. Appl. Mech. Mater. https://doi.org/10.4028/www.scientific.net/AMM.353-356.3511

Yan, B., Janowicz, K., Mai, G., Gao, S., 2017. From ITDL to Place2Vec: Reasoning About Place Type Similarity and Relatedness by Learning Embeddings From Augmented Spatial Contexts, in: Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, SIGSPATIAL '17. ACM, New York, NY, USA, p. 35:1–35:10. https://doi.org/10.1145/3139958.3140054

Yan, Z., Chakraborty, D., Parent, C., Spaccapietra, S., Aberer, K., 2011. SeMiTri: A Framework for Semantic Annotation of Heterogeneous Trajectories [WWW Document]. Proc. 14th Int. Conf. Extending Database Technol. EDBT 2011. URL https://infoscience.epfl.ch/record/163851 (accessed 7.20.20).

Yan, Z., Parent, C., Spaccapietra, S., Chakraborty, D., 2010. A Hybrid Model and Computing Platform for Spatio-semantic Trajectories, in: Aroyo, L., Antoniou, G., Hyvönen, E., ten Teije, A., Stuckenschmidt, H., Cabral, L., Tudorache, T. (Eds.), The Semantic Web: Research and Applications, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 60–75. https://doi.org/10.1007/978-3-642-13486-9_5

Ying, J.J.-C., Chang, Y.-J., Huang, C.-M., Tseng, V.S., 2012. Demographic Prediction Based on User's Mobile Behaviors [WWW Document]. URL /paper/Demographic-Prediction-Based-on-User%27s-Mobile-Ying-Chang/dcbb5c1c469f00fac4b36b99975aa0f3c8719839 (accessed 7.22.20).

Ying, J.J.-C., Lu, E.H.-C., Lee, W.-C., Weng, T.-C., Tseng, V.S., 2010. Mining user similarity from semantic trajectories, in: Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks, LBSN '10. Association for Computing Machinery, San Jose, California, pp. 19–26. https://doi.org/10.1145/1867699.1867703

Zhang, J., Li, X., Yao, Y., Hong, Y., He, J., Jiang, Z., Sun, J., 2020. The Traj2Vec model to quantify residents' spatial trajectories and estimate the proportions of urban land-use types. Int. J. Geogr. Inf. Sci. 0, 1–19. https://doi.org/10.1080/13658816.2020.1726923

Zheng, W., Huang, X., Li, Y., 2017. Understanding the tourist mobility using GPS: Where is the next place? Tour. Manag. 59, 267–280. https://doi.org/10.1016/j.tourman.2016.08.009

Zheng, Y., Zhang, L., Xie, X., Ma, W.-Y., 2009. Mining interesting locations and travel sequences from GPS trajectories, in: Proceedings of the 18th International Conference on World Wide Web, WWW '09. Association for Computing Machinery, Madrid, Spain, pp. 791–800. https://doi.org/10.1145/1526709.1526816

12 Appendix

A. Lists of tags to filter OSM

Land use polygons:

Include when key with any value: "landuse", "natural"

Polygons:

Not to include with key: 'Amenity': "fountain", "atm", "clock", "recycling", "vending_machine", "bench", "water_point", "waste_transfer_station", "waste_disposal", "waste_basket", "post_box", "phone_booth", "parking_space", "parking_entrance", "parking", "motorcycle_parking","fuel", "bicycle_parking", "drinking_water"

Include when key with any value: "amenity", "building", "craft", "historic", "leisure", "military", "office", "shop", "sport", "tourism"

Include when 'other_tag' with the value: 'Amenity', Sport', 'Building', 'Craft', 'Historic', 'Leisure', 'Military', 'Office', 'Shop', 'Tourism', 'Waterway', 'Public_transport', '"Emergency"=>"Ambulance_station"'

***Points*:**
Not to include with key: 'other_tags: '"amenity"=>"drinking_water"', '"amenity"=>"bicycle_parking"', '"natural"=>"tree"', '"natural"=>"tree_row"', '"amenity"=>"fuel"', '"amenity"=>"grit_bin"', '"amenity"=>"motorcycle_parking"', '"amenity"=>"parking"', '"amenity"=>"parking_entrance"', '"amenity"=>"parking_space"', '"amenity"=>"atm"', '"amenity"=>"fountain"', '"amenity"=>"phone_booth"', '"amenity"=>"post_box"', '"amenity"=>"waste_basket"', '"amenity"=>"waste_disposal"', '"amenity"=>"waste_transfer_station"', '"amenity"=>"watering_place"', '"amenity"=>"water_point"', '"amenity"=>"bench"', '"amenity"=>"vending_machine"', '"amenity"=>"recycling"', '"amenity"=>"clock"', '"information"=>"quidepost"'

To include with key: 'other_tags : 'amenity', 'sport', 'building', 'craft', 'historic', 'leisure', 'military', 'office', 'shop', 'tourism', 'waterway', 'public_transport', '"emergency"=>"ambulance_station"', 'natural'

B. Hierarchy of the OSM tags

- Keys in column 'other_tags': ['building', 'garden']
- Tags in column 'other_tags': ['emergency', 'military', 'waterway', 'natural', "historic", 'public_transport', 'tourism', 'leisure', 'amenity', 'craft', 'office', 'shop', 'sport',]
- Column names as keys: ["building", "landuse", "natural", "military", "historic", "tourism", "leisure", "amenity", "craft", "office", "shop", "sport"] *

*Note: as the hierarchy substitutes the information if more relevant information is available, the more important tags are towards the end.

C. 'Common sense' clusters

- **'Green/blue-space'**: ['leisure park','leisure nature_reserve', 'landuse forest', 'landuse meadow', 'natural water', 'leisure playground'],
- **'Exercise'**: ['amenity swimming', 'sport swimming', 'leisure golf_course', 'leisure sports_centre', 'landuse winter_sports', 'sport tennis', 'sport gymnastics'],
- **'Education'**: ['amenity school', 'amenity kindergarten'],
- **'Residential/Building'**: ['landuse residential','building', 'building apartments', 'building farm'],
- **'Health'**: ['amenity hospital','amenity social_facility', 'amenity pharmacy', 'amenity retirement_home', 'amenity doctors'],
- **'Retail'**: ['shop kiosk', 'shop bakery', 'landuse commercial', 'landuse retail', 'shop chemist', 'amenity post_office', 'shop supermarket', 'shop clothes', 'shop electronics', 'shop hairdresser'],
- **'Eating out'**: ['amenity cafe', 'amenity restaurant', 'amenity fast_food', 'amenity bar'],

- **'Tourism'**: ['tourism guest_house','tourism information', 'tourism artwork', 'tourism hotel', 'tourism viewpoint'],
- **'Public transport'**: ['public_transport stop_position', 'public_transport platform', 'landuse railway', 'public_transport station'],
- **'Micellaneous'**: ['landuse allotments','landuse industrial', 'amenity place_of_worship', 'amenity bank', 'leisure common', 'amenity car_sharing', 'landuse cemetery', 'landuse farmyard']

D. Clusters included after omitting too skewed clusters

30-clusters	20-clusters	14-clusters	8-clusters
19	0	0	0
15	1	1	2
24	3	2	3
23	4	4	4
8	5	6	5
1	8	7	7
5	11	8	
0	14	9	
7	15	11	
13	16	12	
16	19		
11			

E. All the clusters and the tags

30-clusters	20-clusters	14-clusters	8-clusters
0 amenity fast_food 0 landuse industrial 0 amenity pharmacy 0 leisure playground 0 sport tennis 0 shop electronics 1 landuse forest 1 tourism museum 2 amenity place_of_worship	0 building apartments 0 landuse commercial 0 landuse winter_sports 1 amenity social_facility 1 amenity hospital 1 tourism guest_house 1 shop bakery 1 landuse railway 1 amenity car_sharing 1 building residential	0 amenity social_facility 0 amenity hospital 0 tourism guest_house 0 landuse allotments 0 shop clothes 0 shop bakery 0 shop chemist	0 amenity restaurant 0 public_transport platform 0 amenity school 0 leisure park 0 shop kiosk 0 amenity fast_food 0 sport swimming 0 landuse industrial 0 shop supermarket

2 natural water 3 landuse commercial landuse 3 winter_sports 4 tourism hotel amenity 5 social_facility 5 amenity hospital 5 shop bakery 5 landuse railway 6 landuse allotments 6 amenity doctors 7 amenity cafe 7 sport gymnastics 7 tourism viewpoint public_transport 8 platform 8 leisure park 8 landuse meadow 9 amenity post_office 10 leisure golf_course amenity 11 kindergarten 11 shop hairdresser public_transport 11 station 12 amenity car_sharing 12 building residential 13 shop supermarket 13 landuse cemetery amenity 14 retirement_home 15 building 16 landuse retail 16 amenity bank 16 landuse farmyard 17 building farm leisure 18 sports_centre 19 landuse residential 20 building apartments 21 sport swimming	2 leisure golf_course amenity 2 retirement_home 3 amenity restaurant 3 amenity school 4 landuse forest 4 tourism museum 5 amenity cafe 5 sport gymnastics amenity 5 kindergarten 5 tourism viewpoint 5 shop hairdresser public_transport 5 station 6 shop clothes 6 amenity bar 7 landuse allotments 7 shop chemist 7 leisure common 7 amenity doctors public_transport 8 platform 8 leisure park 8 landuse meadow 9 amenity post_office 10 sport swimming leisure 11 nature_reserve 12 shop kiosk 12 tourism hotel amenity 13 place_of_worship 13 natural water 14 shop supermarket 14 landuse retail 14 amenity bank 14 landuse cemetery 14 landuse farmyard 15 building 16 amenity fast_food 16 landuse industrial	0 landuse railway 0 leisure common 0 amenity bar amenity 0 car_sharing 0 building residential 0 amenity doctors public_transport 1 platform 1 leisure park 1 shop kiosk 1 amenity fast_food 1 landuse industrial 1 landuse meadow 1 amenity pharmacy 1 leisure playground 1 tourism hotel 1 sport tennis 1 shop electronics 2 amenity restaurant 2 amenity school 2 amenity post_office 3 building farm leisure 3 sports_centre 4 landuse forest 4 tourism museum 5 leisure golf_course amenity 5 retirement_home 6 shop supermarket 6 landuse retail 6 amenity bank 6 landuse cemetery 6 landuse farmyard 7 building 8 building apartments 8 landuse commercial landuse 8 winter_sports 9 landuse residential 10 sport swimming	0 landuse retail 0 landuse meadow 0 amenity post_office 0 amenity bank 0 amenity pharmacy 0 leisure playground 0 tourism hotel 0 sport tennis 0 shop electronics 0 landuse cemetery 0 landuse farmyard 1 building farm 1 leisure golf_course leisure 1 sports_centre amenity 1 retirement_home 2 landuse forest amenity 2 social_facility 2 amenity cafe 2 amenity hospital tourism 2 guest_house 2 landuse allotments 2 shop clothes 2 shop bakery 2 shop chemist 2 landuse railway 2 sport gymnastics 2 leisure common amenity 2 kindergarten 2 amenity bar amenity 2 car_sharing 2 tourism museum 2 tourism viewpoint 2 building residential 2 shop hairdresser public_transport 2 station 2 amenity doctors

22	tourism guest_house	16	amenity pharmacy	11	leisure nature_reserve	3	building apartments
23	leisure nature_reserve	16	leisure playground	12	amenity cafe	3	landuse commercial
24	amenity restaurant	16	sport tennis	12	sport gymnastics	3	landuse winter_sports
25	amenity school	16	shop electronics	12	amenity kindergarten	4	landuse residential
26	shop clothes	17	building farm	12	tourism viewpoint	5	leisure nature_reserve
27	amenity bar	18	leisure sports_centre	12	shop hairdresser	6	amenity place_of_worship
28	shop chemist	19	landuse residential	12	public_transport station	6	natural water
28	leisure common			13	amenity place_of_worship	7	building
29	shop kiosk			13	natural water		

F. List of abbreviations

OSM: OpenStreetMap
POI: Point of Interest
HMM: Hidden Markov Model
POSMIT: Probability of Stops and Moves in Trajectories -method
SE Clustering: Semantic Episode Clustering
NN: Nearest Neighbour -approach
SF-12: short-form health survey
SF-12_01: SF-12 question of self-assessed health
ADS: Depressive Symptoms Score
hd: a spatial stop variance parameter (see: POSMIT)
hi: index search bandwidth parameter (see: POSMIT)
CBOW : Continuous Bag of Words

G. Pseudo code

Semantic enrichment

```
FOR person IN participants:
  FOR Estop IN Tst:
          FOR Pstop IN Pstops:
                  NN point POI
                  Intersect poly POI
                  Intersect land use poly POI
          ENDFOR
          Point POI = MAX((COUNT Point POIs))
          Poly POI = MAX((COUNT Poly POIs))
          Land use POI = MAX(COUNT(Land use POIs))
          IF (MED_DIST(Point POI) < 15 AND COUNT(Point POI) > 180 AND COUNT(Point POI) > 0.5*COUNT(Poly POI)):
                  Esem = Point POI
          ELSEIF (COUNT(Point POI) < COUNT(Poly POI)):
                  Esem = Poly POI
          ELSEIF (MED_DIST(Point POI) < 30 AND COUNT(Point POI) > COUNT(Poly POI)*0.9)
                  POEsemI = Point POI
          ELSEIF (Poly POI == NA AND Land use POI == NA):
                  Esem = Point POI
          ELSEIF (Point POI == NA AND Land use POI == NA):
                  Esem = Poly POI
          ELSEIF (COUNT(Poly POI) > COUNT(Land use POI)*0.5)
                  Esem = Poly POI
          ELSEIF (Esem != Poly POI AND Esem != Point POI)
                  Esem = Land use POI
          RETURN(Esem)
  ENDFOR
  Tsem = CONCAT(Esem)
  RETURN(Tsem)
ENDFOR
```

October 9, 2020, Zurich, Switzerland R. Välimäki

Personal declaration: I hereby declare that the submitted Thesis is the result of my own, independent work. All external sources are explicitly acknowledged in the Thesis

Reetta Välimäki		Munich	9/10/2020
Name	Signature	Place	Date