Measuring activity-based social segregation using public transport smart card data

Kolkowski, Lukas; Cats, Oded; Dixit, Malvika; Verma, Trivik; Jenelius, Erik; Cebecauer, Matej; Rubensson, Isak Jarlebring

DOI
10.1016/j.jtrangeo.2023.103642

Publication date
2023

Document Version
Final published version

Published in
Journal of Transport Geography

Citation (APA)
Kolkowski, L., Cats, O., Dixit, M., Verma, T., Jenelius, E., Cebecauer, M., & Rubensson, I. J. (2023). Measuring activity-based social segregation using public transport smart card data. *Journal of Transport Geography*, *110*, Article 103642. https://doi.org/10.1016/j.jtrangeo.2023.103642

Journal of Transport Geography 110 (2023) 103642

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo

Measuring activity-based social segregation using public transport smart card data

Lukas Kolkowski [a], Oded Cats [a,b,*], Malvika Dixit [a], Trivik Verma [c], Erik Jenelius [b], Matej Cebecauer [b], Isak Jarlebring Rubensson [d]

[a] *Department of Transport and Planning, Delft University of Technology, The Netherlands*
[b] *KTH Royal Institute of Technology, Division of Transport Planning, Sweden*
[c] *Delft University of Technology, Faculty of Technology, Policy and Management, The Netherlands*
[d] *Trafikförvaltningen (Public Transport Administration), Region Stockholm, Stockholm, Sweden*

ARTICLE INFO

Keywords:
Social segregation
Public transport
Ex-post transport appraisal
Smart card data

ABSTRACT

While social segregation is often assessed using static data concerning residential areas, the extent to which people with diverse background travel to the same destinations may offer an additional perspective on the extent of urban segregation. This study further contributes to the measurement of activity-based social segregation between multiple groups using public transport smart card data. In particular, social segregation is quantified using the ordinal information theory index to measure the income group mix at public transport journey destination zones. The method is applied to the public transport smart card data of Stockholm County, Sweden. Applying the index on 2017–2020 data sets for a selected week, shows significant differences between income groups' segregation along the radial public transport corridors following the opening of a major rail project in the summer of 2017. The overall slight decrease in segregation over the years can be linked to declining segregation in the city center as a travel destination and its public transport hubs. Increasing zonal segregation is observed in suburban and rural zones with commuter train stations. This method helps to quantify social segregation, enriching the analysis of urban segregation and can aid in evaluating policies based on the dynamics of social life.

1. Introduction

Social segregation, defined as the uneven spatial distribution of social groups according to Le Roux et al. (2017), is a socio-geographical phenomenon that refers to the lack of mixing amongst social groups, often regarded as a symptom of inequality (Yenn, 2018). Social segregation may lead to disparities in essential living conditions and vice versa (Leonard, 1987; Acevedo-Garcia and Lochner, 2003; Marques, 2012). Inequality and segregation potentially form a reinforcing, vicious circle (Van Ham et al., 2018; Nieuwenhuis et al., 2020). It is therefore paramount to measure the extent of segregation, and thereby enable the assessment of the impacts of alternative interventions and policies thereon.

Spatial segregation of social groups is conventionally measured using segregation indices applied mostly on residential socioeconomic data (Bischoff and Reardon, 2014), i.e. static data reflecting the social mix in the direct vicinity of one's place of residence. While social-demographic data on income, education and housing as well as spatial distance between groups are key drivers of segregation (Tan et al., 2019; United Nations, 2020), considering those only in relation to residential locations arguably offers a limited view of the extent and patterns that characterise segregation within a given geographical area. As reported by Tóth et al. (2021) and Chetty et al. (2022), exposure to and participation in social networks can reinforce inequalities. Given the societal relevance of social segregation, it is of utmost importance to go beyond static measures in order to better reflect the extent to which people from different backgrounds are likely to encounter each other.

Recent studies utilize mobility data to measure segregation using daily travel behavior patterns. We refer to such an undertaking as the analysis of activity-based segregation since travel traces provide insights into activity locations. Past studies that have undertaken such an analysis found significant differences of segregation on weekdays and weekends as well as at different times of the day (Le Roux et al., 2017; Park and Kwan, 2018). Key findings state that mobility usually mitigates social segregation of minorities (Ta et al., 2021). Further, work-related activities are found to reduce levels of segregation (Ellis et al., 2004).

* Corresponding author at: Department of Transport and Planning, Delft University of Technology, The Netherlands.
E-mail address: o.cats@tudelft.nl (O. Cats).

https://doi.org/10.1016/j.jtrangeo.2023.103642
Received 11 February 2023; Received in revised form 23 June 2023; Accepted 23 June 2023
Available online 1 July 2023

Often activity-based social segregation studies rely on self-reported travel diary data or mobile phone data (Farber et al., 2015; Järv et al., 2015) which can induce accuracy, privacy, and availability issues and related biases. Moreover, these data sources have low penetration and population coverage, especially amongst less well-off segments of the population.

Alternatively, data concerning human mobility can be collected by means of passive mobility-specific data collection such as automated fare collection (AFC) schemes in public transport systems. The latter provide unprecedented large datasets of real transactions, i.e., observed mobility traces (Utsunomiya et al., 2006) that allow the measurement of activity-based segregation, albeit limited to public transportation journeys. A couple of studies analysed the travel patterns of specific user groups using public transport smart card data with few measuring the resulting segregation levels. Based on fare reduction for children, seniors and passengers with disabilities registered for smart cards, Abbasi et al. (2021) were able to extract social characteristics to form social groups and measure segregation levels of the respective group compared to the regular fare group. Zhang et al. (2021) had similar options extracting elderly, low-income groups and passengers with disabilities to determine their significantly different activity spaces.

For many transport authorities worldwide, this kind of personal information is partially collected but not made available per travel card holder since extracting it would raise data privacy concerns (Clarke, 2001; Agard et al., 2006). As a result, socioeconomic data often cannot be retrieved directly from smart cards, see Pelletier et al. (2011) and Cats and Ferranti (2022). Moreover, even richly equipped smart cards usually do not contain the desired socio-economic information that is relevant for the analysis of social segregation. It is therefore imperative to develop means to connect social and mobility data to facilitate the analysis of activity-based social segregation based on smart card data traces.

Limited access to transport results in lesser access to essential amenities and opportunities to participate both socially and economically (Lucas, 2011). Transport disadvantage is strongly correlated to social exclusion as found by studies such as Church et al. (2000), Hurni (2006) and Delbosc and Currie (2011). Public transport can potentiality reduce activity-related segregation by offering an affordable means of transport and the analysis thereof will also enable the empirical measurement of the impacts of different interventions and policies.

The potential alleviating effect of transport interventions on activity-based segregation levels can happen within a relatively short time span due to their direct impact on changes in activity (locations) and the resulting travel patterns. In contrast, patterns related to residential segregation - driven by long-term residential location choices as opposed trip generation and distribution choices - are slow to change and the analysis thereof requires a considerably longer time span.

Our study addresses two main research questions. The first question we pose is: How can multi-group activity-based social segregation be measured using large-scale disaggregated mobility data? To this end, we propose a technique for quantifying activity-based segregation by connecting public transport user's mobility patterns and socioeconomic data, thereby making a methodological contribution. The second question we pose is: How does the opening of a major infrastructure project affect activity-based segregation patterns? We conduct an ex-post transport appraisal from the perspective of social segregation. The proposed analysis approach is applied to the case of Stockholm, Sweden and is set into context with the 'Citybanan' commuter train tunnel for the period of 2017–2020, thereby making a substantive contribution. We find that the so-called segregation paradox previously observed in the context of residential segregation (Kovács and Szabó, 2015) may also occur in the context of activity-based segregation.

The remainder of this paper is organised as follows. In the following section we review the literature on social segregation and outline the differences between residential and activity-based segregation (Section 2). We then present our method for measuring activity-based social segregation by enriching mobility data with socioeconomic variables (Section 3). Next, we discuss the application of the method to public transport smart card data from Stockholm, Sweden (Section 4), and present the results thereof (Section 5). We demonstrate the potential ex-post assessment for the case of a rail network investment. We discuss the results and conclude with the key findings and suggestions for further research in Section 6.

2. Literature review

Caused by lack of access to resources, necessities or infrastructure, social segregation is often measured at the place of residence, so-called residential segregation. Relating to movements in daily life, several recent studies have also considered an activity-based approach. This section is dedicated to reviewing literature on measuring residential (2.1) and activity-based (2.2) segregation.

2.1. Residential segregation

The traditional outlook towards segregation is related to peoples' residential location. Within this framing, where people live and how they experience neighbourhoods explains differences in the outcomes of segregation based on ethnic identities, income profiles or the quality of built-environment. Yet, most people do not lead static lives within residential spaces. People move about for their daily activities, work and leisure, and in doing so mix with groups from different local geographies (Kwan, 2013). It is imperative then to understand segregation as a fluid concept in time and expand the lens through which we look at it to not only account for disparities in space but also in activities of people and how they might experience social isolation or exclusion through time. As Kwan (2013) argues that geographic context is important in studying segregation, the author also shows how incorporating human mobility in our analysis helps in discerning the complex and more comprehensive reality of segregation while still retaining the geographic context of people's locations in time.

Most commonly, segregation is explored using non-spatial aggregated data that suffer from significant problems. For example, in these methods, spatial relationships among areal units (neighbourhood selection at the level of data availability) are not accounted for and how those units are discerned impacts the measures of segregation (Openshaw, 1984). Some other researchers have been using spatially aggregated data at the place of residence, an essential life dimension and a commonly available data source through secondary sources like administrative census datasets. Using such static data, social segregation has been found to correlate with socio-demographic characteristics. Past studies on residential social segregation have considered a variety of socioeconomic variables including education, age, and often income (Bischoff and Reardon, 2014). Others considered housing type, educational level, or ethnicity as the indicating factors for segregation (Ivaniushina et al., 2019; Logan and Burdick-Will, 2016).

Income is found to be a major factor influencing segregation. Disparities in income levels cause inequality in access to infrastructure and amenities (Nicoletti et al., 2022) and therefore lead to less mixing and what is commonly referred to as income segregation. This can viciously affect income inequality and perpetuate more staggering forms of segregation in residential places. The relationship between income inequality and residential segregation of socioeconomic groups is prevalent in multiple European cities (Tammaru et al., 2020). Using household data, studies have found increasing income segregation among black and Hispanic US families (Bischoff and Reardon, 2014) and across 12 European capitals (Musterd et al., 2017). Other approaches use data pertaining to work places, schools, or shopping locations, albeit still considering each of which in isolation (Ellis et al., 2004; Rowe and Lubienski, 2017; Erkip, 2003).

Several authors pointed out that the static character of residential data hinders a dynamic view of social mixing in cities, e.g., Xu et al.

(2019), Park and Kwan (2018). Residential segregation studies do not capture the manifold social interactions a human is experiencing throughout the day (Kwan, 2013; Moro et al., 2021). Residential data only leads to conclusions on one socio-geographical space, namely housing, ignoring others such as work or social activities. These associations are often not well-replicated in other geographical contexts or even show inconsistencies in outcomes associated with segregation (e.g. health, wealth or environmental aspects of one's well-being). Both spatial and non-spatial schools of thought only consider residential segregation, overseeing the value in understanding people's trajectories during the day to present a nuanced picture of their experiences of segregation. Further, authors in Wang et al. (2012) and Wong and Shaw (2011) suggest that socio-spatial segregation is a function of exposure to other groups within the individual activity space and research should examine individuals' actual usage and activity patterns in urban space.

2.2. Activity-based segregation

Activity-based segregation studies mostly use geo-located mobility data such as travel survey, and travel diary data (Farber et al., 2015; Xu et al., 2019; Le Roux et al., 2017). Recent studies have demonstrated how the analysis of activity-based segregation can shed light on the extent to which people are exposed to members of other ethnic groups and how the extent of class-based segregation may vary throughout the day based on data in Xining and Paris, respectively (Tan et al., 2019; Le Roux et al., 2017). In the case of Paris, Le Roux et al. (2017) found that segregation is more pronounced in the night locations (vast majority of which pertain to the place of residence) than during the day (pertaining to activity locations). Similarly, the analysis of Tan et al. (2019) showed that different social groups are exposed to diverse environments in their activity locations even if they tend to live in homogeneous neighbourhoods. Both studies are based on a sample of travel diary data. By using surveys, Ta et al. (2021) conclude that the diversity of activities depends on the amenities of residential neighborhoods while pointing out the segregation suffered by migrants. Zhang et al. (2019) combined GPS tracking and activity diary data with the socioeconomic attributes to show that activity-based segregation is influenced by housing types and differs throughout the week.

Building on this call for shift in both the theoretical lens and analytical methods (Kwan, 2013), there is empirical evidence to suggest that activity-based segregation fundamentally differs from residential segregation. Ellis et al. (2004) found significant spatial differences in work and home-related segregation patterns. In particular, they found considerably lower ethnic segregation at work place than in residential spaces. Silm and Ahas (2014) and Athey et al. (2020) report that *experienced segregation* from activities is significantly less pronounced than residential segregation. Advances in this field in the theoretical and methodological (e.g. multi-contextual segregation (Park and Kwan, 2018)) realm also highlight the value of looking at traditional self-reported travel diary data sets with new and enriched concepts of spatiotemporal dynamics involved.

Although these datasets have more nuanced information, they can be improved to reduce some biases in survey methodologies. Some of these methods are subject to errors and rely on a small sample of the population. In addition, studies relying on self-reported travel diary data require immense efforts and high costs to obtain sufficient data sets. Cheaper and more complete data sources, such as data obtained from GPS, mobile phones or social networks, provide accurate mobility traces but often come with privacy concerns. Observed travel patterns can be derived from established technologies such as GPS and fare-collecting public transport smart cards. The latter, while strictly limited to public transport travel, captures actual travel behavior, especially in urban areas, overcoming the shortcomings associated with traditional passenger data collection (Bagchi and White, 2005; Pelletier et al., 2011). The use of public transport smart card data facilitates the progression from "potential mixing" using residence data to "actual exposure/possible contact" because it indicates, to some extent, how people are brought together. In addition, Le Roux et al. (2017) concluded that crossing residential and activity-based data leads to an improved view on social characteristics of populations.

To perform a social segregation analysis, smart card data needs to be enriched with socioeconomic data. Public transport smart card data has been linked to residential socioeconomic data before, but has not as of yet been used in segregation applications. For example, smart card data was enriched with socio-economic data such as income and car ownership to investigate price elasticities (Kholodov et al., 2021) and quantify the spatial extent of travel patterns performed by members of different user groups (Cats and Ferranti, 2022). In this research, we develop a method to measure activity-based social segregation by enriching mobility data in such a way that it connects to travelers' socioeconomic characteristics.

3. Methodology

Following our research objectives, we formulate general requirements regarding mobility data as well as socioeconomic data. Next, we present the steps required for combining socioeconomic data with mobility data and then perform a segregation measurement. In the following, we describe the sequence of steps, along with the respective requirements.

This sequence of steps (shown in Fig. 1) constitutes a methodological framework for measuring social segregation using large-scale disaggregate mobility data and socioeconomic data. Residential socioeconomic data and related grouping are connected to observed disaggregated mobility data by using the same spatial aggregation, e.g. statistical census zones. Hence, socioeconomic and mobility data should be extracted for the same spatial unit of analysis.

First, we identify the social grouping/cluster that best describes each of the zones in the case study area. This can be done by applying either a deterministic or a probabilistic approach. It is therefore assumed that the spatial units considered in the analysis are sufficiently small for users therein to be considered homogeneous, i.e. within zones variability is small compared to between zones variability.

Next, disaggregate mobility data such as smart card data should enable to distinguish and trace individual travellers, i.e. longitudinal mobility traces with consistent user identification are available. Furthermore, destinations are the key to an activity-based approach of measuring segregation. Therefore, this methodology suggests using observed destination-based mobility data, though is not restricted to it.

In order to later assign a traveller to the residential data, a so-called home zone has to be applied so as to identify the zone where the person is most likely to reside. A home-zone is inferred based on the frequency of each zone serving as the origin for journeys performed by the cardholder throughout each of the years included in our analysis. The most common origin zone for the first trip of the day is considered to be the home zone (Kholodov et al., 2021). Once the socioeconomic data is assigned to the mobility data via the inferred travelers' home zones, different segregation measures can be applied. For more details on the rules used for this assignment, the readers are referred to Sari Aslam et al. (2019).

In a third step social groups are linked to each trips destination zone. Matching is performed by inserting the home zones' information of each traveler in the mobility data set. Each journey is made by one distinctive user. This user has a frequent origin and thus a respective home zone. Via the home zone, the socioeconomic information is connected to each transaction the user makes. By attaching for each transaction the respective social classification we are able to proceed with calculating segregation measures.

Next, segregation measures can be applied to the enriched disaggregated mobility data using an index comparing distributions among zones. Our approach quantifies the extent of heterogeneity (or diversity, in terms of the social variable selected for analysis) observed in a given

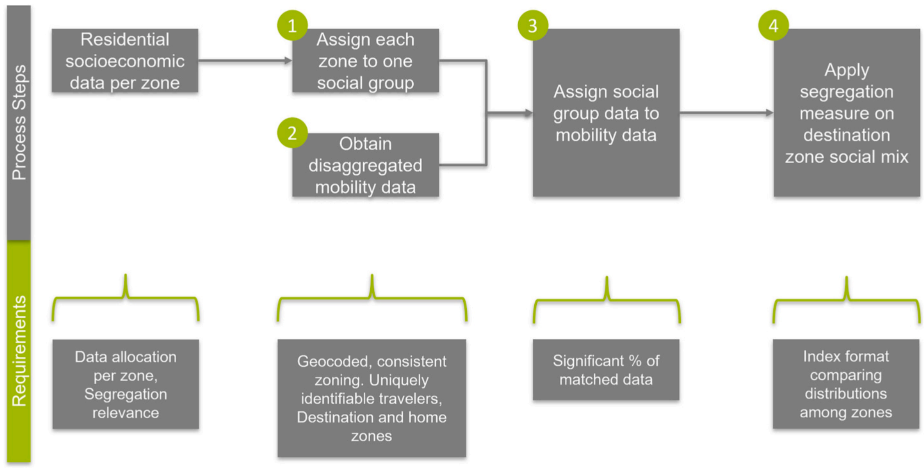

Fig. 1. Framework for measuring social segregation by connecting mobility data and socioeconomic data: The top row presents the sequence of analysis steps and the bottom row displays the corresponding data and analysis requirements.

destination zone by considering the social mix of all travelers to the zone under consideration. Hence, travelers' measured segregation at the journey destination zone, depends on their home zone's social status.

While our approach is not restricted to any specific data type to use, socioeconomic variables that are potentially relevant for describing social segregation are often categorized into predefined groups. For example, share of residents within each income range or the number of residents within each age group. Moreover, certain variables of interest are of ordinal nature, such as the highest level of educational certification. Consequently, these variables have to be treated as ordinal rather than numerical. We therefore adopt the 'ordinal information theory index' (Reardon, 2009) with appropriate adaptations in the following analysis. The ordinal information theory index measures segregation as the ratio of between-category variation to total variation. Although more advanced methods like those based on trajectory-based approaches could provide a more detailed measurement of segregation (Park and Kwan, 2018), they require complete information on the location and duration of all the activities undertaken by an individual. Activity duration are not directly available from smartcard data without making additional assumptions or fusing it with additional data sources such as travel diaries. We therefore choose to calculate segregation using the ordinal information theory index so as to rely solely on the journey destination zone, for which well-established inference methods exist for the analysis of tap-in only smart card data.

The ordinal information theory index is defined as follows

$$\Lambda = \sum_{m\in M}^{M} \frac{t_m}{T\cdot v}(v - v_m) \tag{1}$$

where m is represents a geographical unit such as neighborhoods or census zones, hereafter referred to as zones. m is a member of M which is the set of all zones within the analysis area. t_m is the total population present in zone m and T is the total population within the analysis area. The index is based on the ordinal variation v shown in Eq. (2) which in turn relies on the distribution function f presented in Eq. (3).

$$v = \frac{1}{\|K\| - 1} \sum_{k\in K}^{\|K\|-1} f(c_k) \tag{2}$$

where k is an index of the ordered categories (social groups) and K is the set of all categories. The ratio of $1/[\|K\| - 1]$ is multiplied with the sum of $\|K\| - 1$ values of f, the distribution function defined in Eq. (3). $f(c_k)$ is the cumulative population distribution of the respective ordered category c_k. Using the respective $f(c_k)$, Eq. (2) can be calculated for the analysis area as a whole (v) and as v_m for each individual zone m.

$$f(c) = -[clog_2(c) + (1-c)log_2(1-c)] \tag{3}$$

The closer v_m is to 1 the less homogeneity there is in zone m. Therefore, the value 1 represents the maximum social segregation. Conversely, $v_m = 0$ indicates the maximum amount of homogeneity in zone m, i.e. no segregation. Eq. (1) yields therefore a weighted average term calculated in relation to the overall diversity observed within the case study area.

In order to be able to estimate segregation trends per zone later on, the contribution of each zone to overall segregation is calculated. We distinguish between *absolute* and *weighted* contribution to the segregation index. The absolute contribution is the result of the differences between the total ordinal variation and the zone-specific variation, $v - v_m$. The weighted value corresponds to each zone's contribution to the segregation index calculated by the absolute contribution in relation to the population affected, i.e. $\frac{t_m}{T\cdot v}(v - v_m)$. In other words, the absolute value expresses the contribution before weighting it by the number of passengers affected, while the weighted value accounts for the number of passengers affected relative to the overall number of passengers.

By tracking the mobility of travellers over time and comparing similar time spans, it is possible to measure the extent of segregation observed for different time periods. The ordinal information theory index is hereafter referred to as the *segregation index*. Since it allows for calculating the contributions to the segregation index Λ at the zonal level, the evolution of segregation can be measured even for a single zone. This is done by monitoring the contribution of each zone to the sum and thus the index - the segregation contribution. It allows thus to observe the evolution of each zone's social mix of travelers over time. In the following section we demonstrate the value of this approach.

4. Case study: Stockholm, Sweden

The analysis steps described in Fig. 1 are applied to the multi-modal public transport system of Stockholm County, Sweden.

4.1. Residential segregation

Segregation has been a growing concern and subject of public and political debate in Sweden. Across Swedish cities, ethnic concentrations correlate with concentrated poverty (Malmberg and Clark, 2021). From previous studies based on register data, a negative correlation is observed between income and non-Swedish ethnicity in Stockholm (Harsman, 2006; Tammaru et al., 2020). Segregation in Stockholm is mostly connected to the spatial clustering of residents' with specific ethnic background and income levels in distinctive residential zones

(Andersson and Kährik, 2015).

The national land allocation system which aims at mixed housing could have potentially played a mitigating role, by mixing municipally controlled developers in combination with land ownership (Caesar and Kopsch, 2018). Nevertheless, the capital city, Stockholm is found to be an increasingly socially segregated city by an array of studies (Andersson and Kährik, 2015; Nielsen and Hennerdal, 2017; Haandrikman et al., 2021). Hedin et al. (2012) concluded that there has been a "growth of super-gentrification and low-income filtering" in Sweden between 1986–2001. In recent years, especially low-income groups seem to be segregated in urban outskirts (Grundström and Molina, 2016). While in the case of Stockholm gentrification took place mostly in the northern and eastern parts of the metropolitan area, the low-income filtering tended towards the northwest and southwest corridors. Consequently, Stockholm has more residential poverty segregation than other European metropolises such as Amsterdam, Brussels, Copenhagen and Oslo (Haandrikman et al., 2021).

As discussed in Section 2 income greatly influences social segregation. As evident from the discussion above, this is not least the case for Stockholm.

4.2. Public transport system and the Citybanan project

The capital's region is home to 2.4 million inhabitants and has a well-developed public transport network integrated into one fare-collecting smart card. Public transport in Stockholm has a high modal share, and both low and high-income residents use it. The average modal share for all days and all journeys is 30 percent. And although low-income residents are over-represented as users, high-income users still do a fair amount of traveling by public transport. The modal share for public transport is 26 percent for travelers from households in the top 30 percent of the income distribution (Johansson, 2020).

The Citybanan, a new commuter train tunnel in Stockholm's city center, opened in July 2017. It was build to relieve a national and regional bottleneck in the corridor of the inner Stockholm City. This major network change led to the separation of commuter trains and regional/national train tracks in the inner-city, see Fig. 2.

Since its opening in July 2017, the public transport of Stockholm County changed accordingly in the adaptation of the commuter train lines, so-called Pendeltåg. The project was considered to both significantly enhance punctuality and increase traffic throughput for urban, regional, and even national transport and thereby increase the level of service for travelers. Part of the investment motivation was to reduce segregation and improve accessibility for outer suburbs.

4.3. Data processing

Smart card data. In this study we consider a representative week, the fifth week of the years 2017–2020. In 2020, 8.45 million journeys were conducted between Jan 27 and Feb 3. In July 2017, the new commuter train corridor Citybanan was inaugurated, which allows for increased frequency and accessibility for connections between some of the suburbs and neighboring municipalities and Stockholm inner-city. We demonstrate the applicability of our method for the analysis of potential changes in segregation by investigating data from the fifth week of the year for all years between 2017 and 2020.

The fare scheme in Stockholm is flat and the automated fare collection system involves tapping-in only. All smart card transactions were processed to generate a database with a detailed travel diary for each card-holder for each of the weeks included in the analysis. The construction of this data set involves the sequential implementation of a trip destination inference algorithm and a transfer inference algorithm, the details of which are provided in Cats and Ferranti (2022). Next, we are interested in identifying the most likely census home-zone for each card-holder as described in Section 3.

Connecting mobility and social data. Given the relevance of income segregation in Stockholm, we examine income data per zone to form socio-economic groups. The 1300 so-called DeSo (Demographic statistics areas) census zones have populations ranging from 700 to 2700, i.e. each of which accounting for 0.03–0.12 percent of the Stockholm county population. For each zone, the median annual income and the share of residents within each income quantile of all 20 + years residents is available from the 2017 Swedish income and tax register (SCB, 2017). To connect the residential-based income and tax register data to the public transport smart card data, a vector containing the share of all residents associated with each of the income quantiles is assigned to each home zone. The socioeconomic data is connected to each smart card transaction through the zone ID. The segregation index is then calculated for each DeSo zone by considering the income characteristics of the origin zone of all passenger journeys destined to this

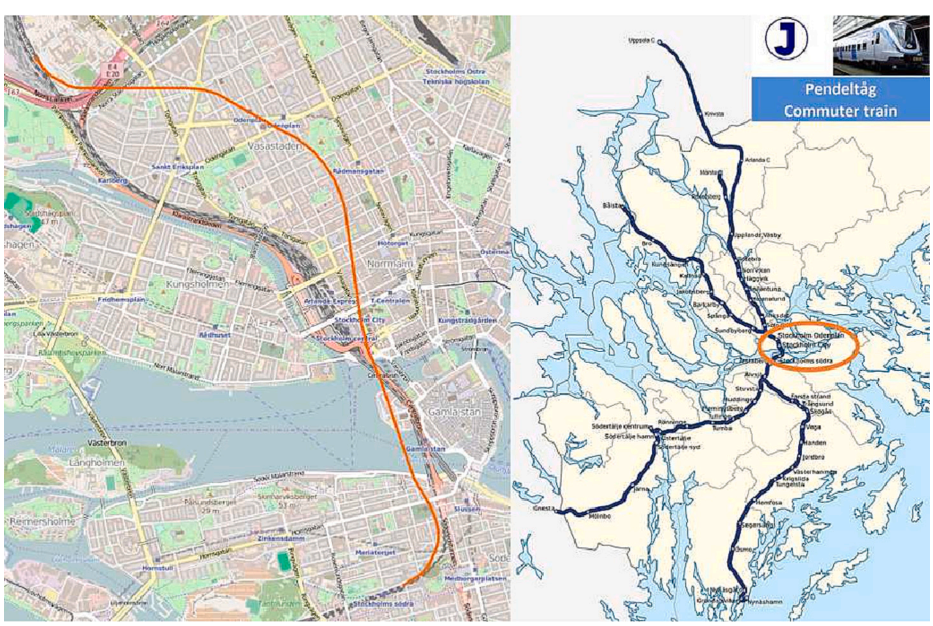

Fig. 2. Left: Citybanan tunnel Stockholm (regional rail tracks: black-white lines, commuter train tunnel: orange), Ellgaard (2009). right: Commuter train bottleneck in Stockholm County, Frohne (2014).

zone. In the following, we measure the potential exposure in each destination zone and the resulting segregation index on a daily basis for each of the days included in our dataset.

5. Results: Activity-based income segregation

5.1. System-wide segregation level per year

Our results highlight that the activity-based social segregation in residential areas reduced over the years as shown by the index score (see Eq. (1)) range between an average of 0.1923 in 2017 and 0.1888 in 2020. Lower segregation levels are recorded for the intermediate years, 0.1877 (for 2018, a decrease of 2.4% compared to 2017) and 0.1856 (for 2019, 3.5% lower compared to 2017), respectively.

5.2. Variations over days of the week

When analysing how the segregation index varies for different days of the week we find consistent patterns for all four years included in our analysis. Fig. 3 presents the average zonal segregation index, Λ (Eq. (1)), for each day-of-the-week for each of the years 2017–2020. The results indicate that people mix to a similar extent on Monday to Thursday. Conversely, the lowest segregation level is observed on Fridays and Saturdays. These can be related to the combination of work, leisure, and shopping activities. In contrast, on Sundays, travelers mix less and experience more segregation as the index is higher than on other days, as residents are less likely to travel to more diverse zones, for example for work purposes. These results are consistent with the findings of activity-based analysis in the United States, which assert that work-related activities reduce segregation (Ellis et al., 2004).

Additionally, Fig. 4 displays the average weighted and “absolute” segregation contribution, respectively, of each zone in 2017 (see details in the Methodology section). These are average values across all days of week 5. As can be observed, the weighted segregation contribution is highest in central zones and suburban centers. In contrast, the absolute segregation is highest for peripheral zones (see insert at the top-left of the figure). Popular destinations such as the city center zones and key commercial destinations have a low absolute difference but a high contribution once accounting for the large volumes of passengers visiting these zones. These maps allow policy makers to identify parts of the metropolitan area which are most likely to be visited by travellers from different backgrounds (in this case income level) as well as those ares which exhibit low diversity in terms of visitors’ social mix-up.

5.3. Ex-post appraisal of changes in segregation

The analysis of spatial disparities in activity-based segregation patterns can be further utilised to investigate how those have evolved over time. By taking the difference between weighted segregation contributions for selected years, it can be examined whether a specific zone contributes to an increase or a decrease in the segregation index value. Thereby, the evolution of segregation can be assessed at the zonal level. A negative difference thus indicates a decline in segregation. Conversely, a positive difference shows an increase in contribution to segregation.

As mentioned in Section 4, a major infrastructure project, the Citybanan, was completed in the summer of 2017. The opening of this infrastructure was accompanied by significant changes to the public transport network in the Stockholm County and consequently the resulting accessibility across the region. We investigate year-on-year differences between 2018 and 2017. Changes in observed segregation levels reflect underlying changes in travel demand patterns and the associated income levels.

Decreasing segregation is found in fringes of the city center and suburban zones located in direct proximity to the commuter train stations, and especially along the southwest corridor (Fig. 5). This decreasing trend in segregation levels recorded in central areas is largely sustained in subsequent years (2019 and 2020). At the same time, segregation increased in several suburban and peri-urban areas located along the northwest, southwest and south commuter train corridors which saw ridership increase, mostly by residents of these zones which have a more homogeneous income profile than the region as a whole. However, these zones are associated with significantly fewer visits than the central ones. Overall, zones with segregation reductions outnumber those who saw increases in 2018. As mentioned in the opening of this section, in total, segregation decreases by 2.4% year-on-year.

Urban zones with public transport hubs and primarily commuter train stations such as the newly created Odenplan station display a decreasing trend in segregation values over the years (Fig. 6). Consequently, city center inbound public transport passengers are found to be more income-diverse after the opening of the Citybanan than before.

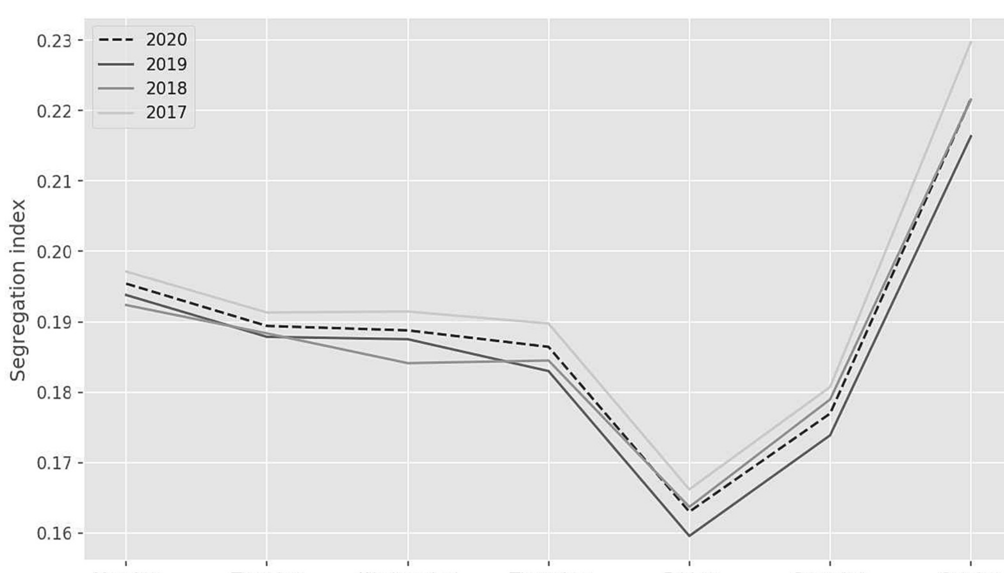

Fig. 3. Segregation index Λ throughout the week. Segregation index Λ ranges within [0,1] with 1 indicating no income group mixing at the destination zone and therefore maximum segregation and 0 indicating equally distributed income groups over all destination zones, thus no segregation.

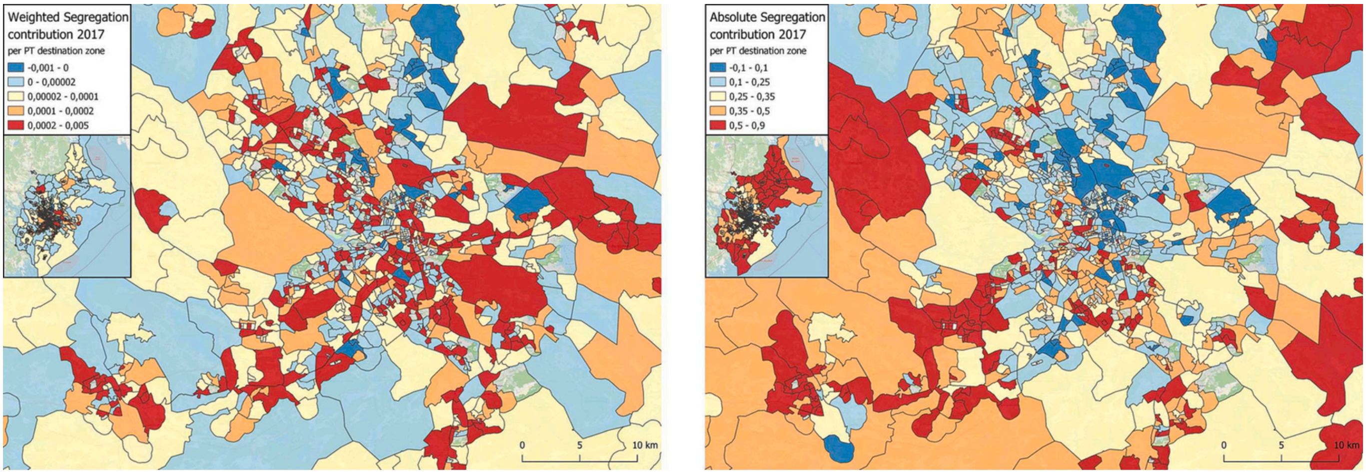

Fig. 4. Weighted (left) and absolute (right) segregation contribution 2017 - each social-demographic zone's arriving public transport passenger mix contribution to the overall segregation index level.

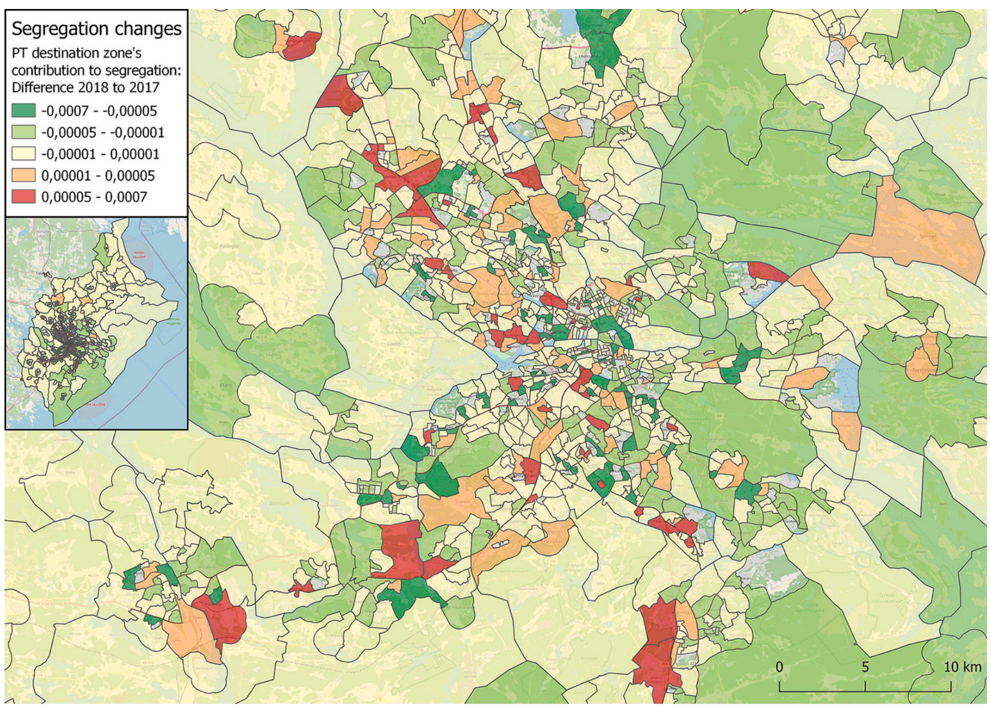

Fig. 5. Zonal segregation year-on-year changes for 2017–2018 - decreasing levels indicate less segregation contribution.

6. Conclusions

We demonstrate how connecting large-scale mobility data to social data can enrich our understanding of social segregation and examine segregation developments in relation to transport or policy changes. Daily or weekly segregation levels help evaluate overall levels and trends. Weighted segregation levels are suitable for analyses in which the relation of zone segregation plays a role. The absolute segregation contribution may be used for detailed, zonal assessment. Particularly for urban planners and policymakers, it could be of interest to measure how social segregation evolves, assessing the impacts of various interventions such as fare scheme change and network developments. In addition, the proposed metric allows comparing segregation levels across cities and regions.

Our application for the case of Stockholm indicates mixed effects of the new Citybanan commuter rail corridor. The majority of zones remaining largely unchanged. City center inbound passengers are found to become more income-diverse, while outbound passengers towards the suburbs have a more uniform income backgrounds, especially when traveling to commuter train stations. Consequently, decreasing segregation is found in fringes of the city center and suburban zones located in direct proximity to the commuter train stations, and especially along the southwest commuter rail corridor. This is presumably attributed to the main change that occurred between the measurement periods, namely the opening of the Citybanan which improved access between mostly less affluent suburbs and municipalities and the central districts of Stockholm. Meanwhile, increasing segregation is found around the public transport hubs along the northwest, southwest and south commuter train corridors. Increasing segregation levels in these suburban and peri-urban zones could be linked to general trends of urbanization and gentrification, as well as the transport disadvantage of low-income groups. These corridors have previously been observed as low-

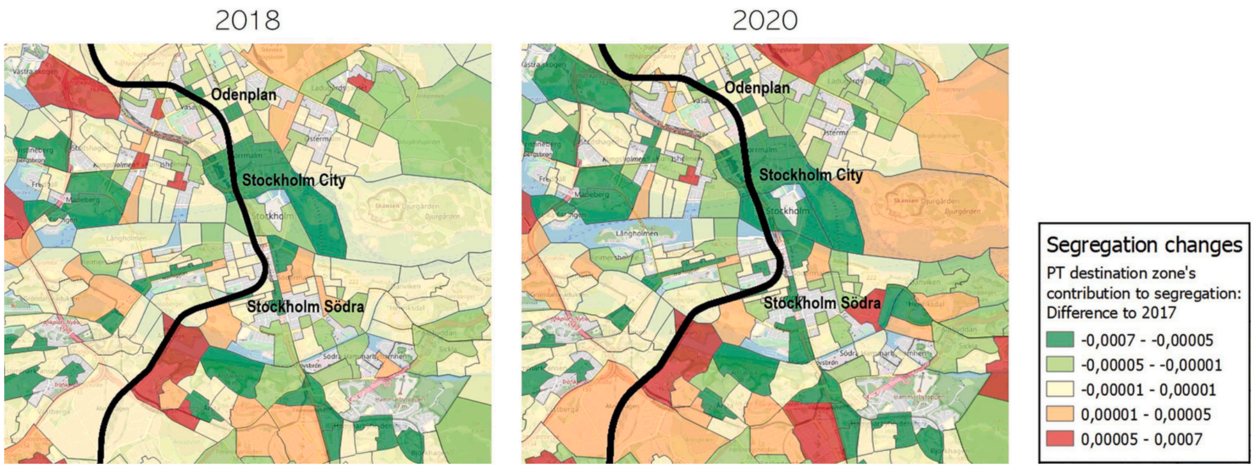

Fig. 6. Stockholm city center segregation changes - Differences in zonal segregation contribution in 2018 and 2020 compared to 2017. Commuter train lines are drawn as black lines.

income areas with increasing residential segregation between 1990 and 2010, partly due to a change in housing policy (Andersson and Turner, 2014; Andersson and Kährik, 2015). The increase in activity-based segregation is in line with these findings and indicates an increase in social segregation in these "pockets of poverty". When examining the effects of urban planning measures in Stockholm, as studied by (Andersson et al., 2010), our activity-related results could be taken into account in addition to the analysis of residential segregation.

Disaggregate demand models such as agent-based transport assignment models allow for the analysis and assessment of such (de-) segregation impacts as part of project appraisal process. This will in turn support decision makers in accounting for segregation effects when considering alternative investments or interventions.

When comparing segregation levels for different days of the week, we find that Fridays and Saturdays are associated with the lowest segregation levels whereas Sundays register a considerably higher level of segregation than other days of the week. This finding resonates with the following statement made in 1963, albeit made in a very different societal context, by Martin Luther King "It is appalling that the most segregated hour of Christian America is 11 o'clock on Sunday morning".

In the context of residential segregation, the so-called segregation paradox is observed when high-income households move into low-income neighbourhoods and temporary mixing of socioeconomic groups may occur (Kovács and Szabó, 2015). A similar phenomenon may occur also in the context of activity-based segregation. For example, if high-income passengers are visiting more frequently areas which have been previously predominantly visited by low-income individuals. This would then yield lower levels of segregation, which is technically correct but not necessarily desired as it could also be an indication of gentrification. While the limited amount of dwelling units and the strong real-estate pricing effects are much more severe in relation to residential choice, the availability of amenities and commerce catering for individuals from different income levels also constitutes a limited resource. Reduced segregation level therefore should not automatically be considered a desirable outcome and a more detailed analysis might be performed to shed additional light on local dynamics.

The proposed method allows measuring and quantifying segregation - as measured in terms of the diversity of income mix-up at travel destinations in this study - and spatio-temporal changes therein. However, it does not shed light on the underlying determinants. Future research may investigate the processes resulting in observed activity-based segregation patterns as well as disentangling the effects that can be linked to specific observed changes. Moreover, segregation is a multi-faced phenomena and the activity-based segregation associated with social variables of interest other than income - such as education level, occupation type, migration background and political views - may also be analysed by adopting the approach employed in this study. Similarly, the analysis can be extended to studying within-day temporal variations in the segregation for different social groups. Even though the size of the zones employed in our analysis are considered sufficiently small to allow for high homogeneity of individuals residing therein, especially considering past empirical findings regarding segregation patterns in Stockholm (Hedin et al., 2012; Andersson and Kährik, 2015; Grundström and Molina, 2016; Nielsen and Hennerdal, 2017; Haandrikman et al., 2021), there could be a future avenue of research linking truly individual social data from, e.g., a travel survey to mobility traces to test if the findings hold when avoiding the risk for ecological bias.

Future study may fuse data collected from various travel modes to establish segregation levels once accounting for all movements and possibly identify whether different modes play distinctive roles in contributing or possibly mitigating the impacts of residential segregation. Further, changes in measured activity-based segregation might be incorporated into a multi-criteria policy analysis and transport investment assessment to reflect related societal objectives. Finally, the proposed method can be applied to analyze social segregation in relation to other dimensions as well, e.g. education level, voting patterns or ethnic background in order to study (social) mobility and the risk for polarisation of different social groups.

CRediT authorship contribution statement

Lukas Kolkowski: Methodology, Data-curation, Formal-analysis, Writing-original-draft, Writing-review-editing. **Oded Cats:** Conceptualization, Methodology, Investigation, Writing-original-draft, Writing-review-editing, Project-administration, Funding-acquisition. **Malvika Dixit:** Supervision, Writing-review-editing. **Trivik Verma:** Supervision, Writing-review-editing. **Erik Jenelius:** Methodology, Supervision, Writing-review-editing. **Matej Cebecauer:** Software, Data-curation. **Isak Jarlebring Rubensson:** Validation, Supervision.

Data availability

The data that has been used is confidential.

Acknowledgment

The authors would like to thank Region Stockholm for sharing the data used in this study. This study is funded by Region Stockholm, project "Unravelling travel demand patterns using Access card data"RS 2019–0499 and "CAPA-CITY: Identifying capacity gaps to support urban and regional development”RS 2022–0210.

References

Abbasi, S., Ko, J., Min, J., 2021. Measuring destination-based segregation through mobility patterns: Application of transport card data. J. Transp. Geogr. 92.

Acevedo-Garcia, D., Lochner, K.A., 2003. Residential segregation and health. Neighborhoods and health, pp. 265–287.

Agard, B., Morency, C., Trépanier, M., 2006. Mining public transport user behaviour from smart card data. IFAC Proc. Vol. 39 (3), 399–404.

Andersson, R., Kährik, A., 2015. Widening gaps: Segregation dynamics during two decades of economic and institutional change in stockholm. In: Socio-Economic Segregation in European Capital Cities. Routledge, New York, pp. 134–155.

Andersson, R., Turner, L.M., 2014. Segregation, gentrification, and residualisation: From public housing to market-driven housing allocation in inner city stockholm. Int. J. Hous. Policy 14, 3–29.

Andersson, R., Bråmå, Åsa, Holmqvist, E., 2010. Counteracting segregation: Swedish policies and experiences. Hous. Stud. 25 (2), 237–256.

Athey, S., Ferguson, B.A., Gentzkow, M., Schmidt, T., 2020. Experienced segregation. Technical report. National Bureau of Economic Research.

Bagchi, M., White, P.R., 2005. The potential of public transport smart card data. Transp. Policy 12, 464–474.

Bischoff, K., Reardon, S.F., 2014. Residential segregation by income, 1970-2009. Diversity and disparities: America enters a new century, 43.

Caesar, C., Kopsch, F., 2018. Municipal land allocations: a key for understanding tenure and social mix patterns in stockholm. Eur. Plann. Stud. 26, 1663–1681.

Cats, O., Ferranti, F., 2022. Unravelling the spatial properties of individual mobility patterns using longitudinal travel data. J. Urb. Mob. 2, 100035.

Cats, O., Ferranti, F., 2022. Voting with one’s feet: Unraveling urban centers attraction using visiting frequency. Cities 127, 103773.

Chetty, R., Jackson, M.O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R.B., Gong, S., Gonzalez, F., Grondin, A., Jacob, M., Johnston, D., Koenen, M., Laguna-Muggenburg, E., Mudekereza, F., Rutter, T., Thor, N., Townsend, W., Zhang, R., Bailey, M., Barberá, P., Bhole, M., Wernerfelt, N., 2022. Social capital ii: determinants of economic connectedness. Nature 608, 122–134.

Church, A., Frost, M., Sullivan, K., 2000. Transport and social exclusion in London. Transp. Policy 7 (3), 195–205.

Clarke, R., 2001. Person location and person tracking-technologies, risks and policy implications. Inf. Technol. People.

Delbosc, A., Currie, G., 2011. Transport problems that matter - social and psychological links to transport disadvantage. J. Transp. Geogr. 19, 170–178.

Ellgaard, H., 2009. Citybanan Stockholm op de kaart. Wikipedia. Retrieved fromhttps://nl.wikipedia.org/wiki/Citybanan_Stockholm#/media/Bestand:Citybanan_karta.jpg on October 11, 2021.

Ellis, M., Wright, R., Parks, V., 2004. Work together, live apart? geographies of racial and ethnic segregation at home and at work. Ann. Assoc. Am. Geogr. 94, 620–637.

Erkip, F., 2003. The shopping mall as an emergent public space in turkey. Environ. Plann. A: Econ. Space 35 (6), 1073–1093.

Farber, S., O’Kelly, M., Miller, H.J., Neutens, T., 2015. Measuring segregation using patterns of daily travel behavior: A social interaction based model of exposure. J. Transp. Geogr. 49, 26–38.

Frohne, E., 2014. Strecke der Pendeltåg (Stockholm). Wikipedia. Retrieved fromhttps://de.wikipedia.org/wiki/Pendelt%C3%A5g_(Stockholm)#/media/Datei:Linjekarta_f%C3%B6r_Stockholms_pendelt%C3%A5g.jpg on October 11, 2021.

Grundström, K., Molina, I., 2016. From folkhem to lifestyle housing in sweden: segregation and urban form, 1930s–2010s. Int. J. Hous. Policy 16, 316–336.

Haandrikman, K., Costa, R., Malmberg, B., Rogne, A.F., Sleutjes, B., 2021. Socio-economic segregation in european cities. a comparative study of brussels, copenhagen, amsterdam, oslo and stockholm. Urb. Geogr.

Harsman, B., 2006. Ethnic diversity and spatial segregation in the stockholm region. Urb. Stud. 43 (8), 1341–1364.

Hedin, K., Clark, E., Lundholm, E., Malmberg, G., 2012. Neoliberalization of housing in sweden: Gentrification, filtering, and social polarization. Ann. Assoc. Am. Geogr. 102, 443–463.

Hurni, A., 2006. Transport and social exclusion in western sydney transport and social exclusion in western sydney.

Ivaniushina, V., Makles, A.M., Schneider, K., Alexandrov, D., 2019. School segregation in st. petersburg–the role of socioeconomic status. Educ. Econ. 27, 166–185.

Johansson, M., 2020. Resvaneundersökning 2019. Trafikförvaltningen. SL 2018-0116.

Järv, O., Müürisepp, K., Ahas, R., Derudder, B., Witlox, F., 2015. Ethnic differences in activity spaces as a characteristic of segregation: A study based on mobile phone usage in tallinn, estonia. Urb. Stud. 52, 2680–2698.

Kholodov, Y., Jenelius, E., Cats, O., van Oort, N., Mouter, N., Cebecauer, M., Vermeulen, A., 2021. Public transport fare elasticities from smartcard data: Evidence from a natural experiment. Transp. Policy 105, 35–43.

Kovács, Z.L.Á.N., Szabó, B., 2015. Urban restructuring and changing patterns of socio-economic segregation in budapest. In: Socio-Economic Segregation in European Capital Cities. Routledge, pp. 262–284.

Kwan, M.P., 2013. Beyond space (as we knew it): Toward temporally integrated geographies of segregation, health, and accessibility: Space-time integration in geography and giscience. Ann. Assoc. Am. Geogr. 103, 1078–1086.

Le Roux, G., Vallée, J., Commenges, H., 2017. Social segregation around the clock in the paris region (france). J. Transp. Geogr. 59, 134–145.

Leonard, J.S., 1987. The interaction of residential segregation and employment discrimination. J. Urb. Econ. 21 (3), 323–346.

Logan, J.R., Burdick-Will, J., 2016. School segregation, charter schools, and access to quality education. J. Urb. Aff. 38, 323–343.

Lucas, K., 2011. Making the connections between transport disadvantage and the social exclusion of low income populations in the tshwane region of south africa. J. Transp. Geogr. 19, 1320–1334.

Malmberg, B., Clark, W.A., 2021. Migration and neighborhood change in sweden: The interaction of ethnic choice and income constraints. Geogr. Anal. 53 (2), 259–282.

Marques, E., 2012. Social networks, segregation and poverty in são paulo. Int. J. Urban Reg. Res. 36, 958–979.

Moro, E., Calacci, D., Dong, X., Pentland, A., 2021. Mobility patterns are associated with experienced income segregation in large us cities. Nat. Commun. 12, 4633.

Musterd, S., Marcińczak, S., van Ham, M., Tammaru, T., 2017. Socioeconomic segregation in european capital cities. increasing separation between poor and rich. Urb. Geogr. 38, 1062–1083.

Nicoletti, L., Sirenko, M., Verma, T., 2022. Disadvantaged communities have lower access to urban infrastructure.

Nielsen, M.M., Hennerdal, P., 2017. Changes in the residential segregation of immigrants in sweden from 1990 to 2012: Using a multi-scalar segregation measure that accounts for the modifiable areal unit problem. Appl. Geogr. 87, 73–84.

Nieuwenhuis, J., Tammaru, T., van Ham, M., Hedman, L., Manley, D., 2020. Does segregation reduce socio-spatial mobility? evidence from four european countries with different inequality and segregation contexts. Urb. Stud. 57, 176–197.

Openshaw, S., 1984. The modifiable areal unit problem. Concepts Tech. Modern Geogr.

Park, Y.M., Kwan, M.-P., 2018. Beyond residential segregation: A spatiotemporal approach to examining multi-contextual segregation. Comput. Environ. Urban Syst. 71, 98–108.

Pelletier, M.P., Trépanier, M., Morency, C., 2011. Smart card data use in public transit: A literature review. Transp. Res. Part C: Emerg. Technol. 19, 557–568.

Reardon, S.F., 2009. Measures of ordinal segregation. In: Occupational and residential segregation. Emerald Group Publishing Limited.

Rowe, E.E., Lubienski, C., 2017. Shopping for schools or shopping for peers: public schools and catchment area segregation. J. Educ. Policy 32 (3), 340–356.

Sari Aslam, N., Cheng, T., Cheshire, J., 2019. A high-precision heuristic model to detect home and work locations from smart card data. Geo-Spat. Inf. Sci. 22, 1–11.

SCB, 2017. SCB:s Open data for DeSO – Demographic Statistical Areas. Retrieved fromhttps://www.geodata.se/geodataportalen/srv/swe/catalog.search;jsessionid=42B5AAC3339638A205A27724ECF960BF#/search?resultType=swe-details&_schema=iso19139*&type=dataset%20or%20series&from=1&to=20 using the explanation fromhttps://www.scb.se/en/services/open-data-api/open-geodata/deso-demographic-statistical-areas/.

Silm, S., Ahas, R., 2014. The temporal variation of ethnic segregation in a city: Evidence from a mobile phone use dataset. Soc. Sci. Res. 47, 30–43.

Ta, N., Kwan, M.-P., Lin, S., Zhu, Q., 2021. The activity space-based segregation of migrants in suburban shanghai. Appl. Geogr. 133, 102499.

Tammaru, T., Marcińczak, S., Aunap, R., van Ham, M., Janssen, H., 2020. Relationship between income inequality and residential segregation of socioeconomic groups. Reg. Stud. 54, 450–461.

Tan, Y., Chai, Y., Chen, Z., 2019. Social-contextual exposure of ethnic groups in urban china: From residential place to activity space. Popul. Space Place 25.

Tóth, G., Wachs, J., Clemente, R.D., Jakobi, Ákos, Ságvári, B., Kertész, J., Lengyel, B., 2021. Inequality is rising where social network segregation interacts with urban topology. Nat. Commun. 12.

United Nations, 2020. World Social Report 2020: inequality in a rapidly changing world. United Nations Department of Economic and Social Affairs. United Nations publication, New York. Sales No. E.20.IV.1.

Utsunomiya, M., Attanucci, J., Wilson, N., 2006. Potential uses of transit smart card registration and transaction data to improve transit planning. Transp. Res. Rec. 1971 (1), 118–126.

Van Ham, M., Tammaru, T., Janssen, H.J., 2018. A multi-level model of vicious circles of socio-economic segregation. Divided cities: understanding intra-urban disparities. OECD, Paris.

Wang, D., Li, F., Chai, Y., 2012. Activity spaces and sociospatial segregation in beijing. Urb. Geogr. 33 (2), 256–277.

Wong, D.W., Shaw, S.L., 2011. Measuring segregation: An activity space approach. J. Geogr. Syst. 13, 127–145.

Xu, Y., Belyi, A., Santi, P., Ratti, C., 2019. Quantifying segregation in an integrated urban physical-social space. J. R. Soc. Interface 16.

Yenn, T.Y., 2018. This is what inequality looks like. Ethos Books.

Zhang, S., Yang, Y., Zhen, F., Lobsang, T., Li, Z., 2021. Understanding the travel behaviors and activity patterns of the vulnerable population using smart card data: An activity space-based approach. J. Transp. Geogr. 90.

Zhang, X., Wang, J., Kwan, M.-P., Chai, Y., 2019. Reside nearby, behave apart? activity-space-based segregation among residents of various types of housing in beijing, china. Cities 88, 166–180.