Characterizing residential segregation in cities using intensity, separation, and scale indicators

Spierenburg, Lucas; van Cranenburgh, Sander; Cats, Oded

DOI
10.1016/j.compenvurbsys.2023.101990

Publication date
2023

Document Version
Final published version

Published in
Computers, Environment and Urban Systems

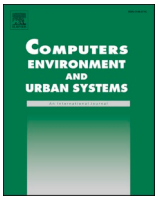
Citation (APA)
Spierenburg, L., van Cranenburgh, S., & Cats, O. (2023). Characterizing residential segregation in cities using intensity, separation, and scale indicators. *Computers, Environment and Urban Systems*, *103*, Article 101990. https://doi.org/10.1016/j.compenvurbsys.2023.101990

Computers, Environment and Urban Systems 103 (2023) 101990

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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



Characterizing residential segregation in cities using intensity, separation, and scale indicators

Lucas Spierenburg [a,*], Sander van Cranenburgh [b], Oded Cats [a]

[a] Transport and Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Netherlands
[b] Transport and Logistics Group, Department of Engineering Systems and Services, Faculty of Technology, Policy, and Management, Delft University of Technology, Netherlands

ARTICLE INFO

Keywords:
Residential segregation
Regionalization
Spatial clustering
Multi-dimensional analysis
Spatial scale

ABSTRACT

This paper proposes a method to characterize residential segregation patterns along three dimensions: intensity, separation, and scale. These dimensions designate respectively the over-representation of a group in segregated regions, the proportion of people from that group living in these regions, and the spatial extent of these regions. We apply the method to all Dutch municipalities, to study segregation along migration background. Our results demonstrate that no single segregation pattern prevails in the Netherlands: Dutch municipalities express very diverse combinations of intensity, separation, and scale. We then explore the relation between segregation patterns and municipality characteristics. We show that segregation intensity and separation associate with municipality size, income inequality, and share of the group of interest in the municipality population. Moreover, the scale of segregation is highly correlated with municipality size, suggesting a proportional relation.

1. Introduction

Residential segregation, here understood as the uneven distribution of social groups in space (Reardon & O'Sullivan, 2004), is a persistent problem in many cities in the world (Tammaru, van Ham, Marcińczak, & Musterd, 2015; Wang, Phillips, Small, & Sampson, 2018). Occurring along one or several social characteristics — such as income, religion, or migration background — segregation often exacerbates inequality across groups in many aspects of people's life, notably in education achievements, well-being, and health condition (Ludwig et al., 2012; Nieuwenhuis, Tammaru, van Ham, Hedman, & Manley, 2020; Owens, 2018; Tóth et al., 2021; Williams & Collins, 2001). Consequently, numerous local and national authorities consider residential segregation a priority issue (the United Nations, 2017; Gemeente Amsterdam, 2017; Ville de Paris, 2021; Mayor of London, 2021). In order to adequately tackle residential segregation, these authorities need tools for characterizing it and for understanding its underlying determinants. A body of literature addresses these two needs, proposing indicators to quantify residential segregation patterns, and relating these patterns to city characteristics such as city size, social mix, or income inequality (Durlauf, 1996; Gordon & Monastiriotis, 2006; Krupka, 2007; Massey & Denton, 1988; Scarpa, 2015; Theil & Finizza, 1971; Veneri, Comandon, Àngel Garcia-López, & Daams, 2021).

When characterizing segregation patterns, policymakers should account for the multi-dimensionality of the phenomenon (Massey & Denton, 1988; Petrović, van Ham, & Manley, 2018). Widely studied dimensions include: the degree to which groups are over-represented in certain regions, the proportion of people living in regions in which their group is over-represented, and the spatial extent of these regions; respectively called segregation intensity, separation, and scale in our paper (Andersson et al., 2018; Clark, Anderson, Östh, & Malmberg, 2015; Fowler, 2016; Lan, Kandt, & Longley, 2020; Olteanu, Randon-Furling, & Clark, 2019; Östh, Clark, & Malmberg, 2015; Reardon & O'Sullivan, 2004). Capturing all these dimensions independently is crucial for two reasons. First, they contribute separately to the segregation experienced by the population. For example, for the same intensity of residential segregation, larger spatial scale may result in higher segregation at school, as distance plays an important role in the allocation of children to schools (Gramberg, 1998; Taylor & Gorard, 2001). Second, policies addressing residential segregation may not necessarily affect all dimensions equally.

Several studies have introduced frameworks capturing segregation intensity, separation, and scale, but they all share the shortcoming of measuring separation and scale indirectly, i.e. as a function of the

* Corresponding author at: Stevinweg 1 2628 CN, Delft, the Netherlands.
E-mail address: l.j.spierenburg@tudelft.nl (L. Spierenburg).

https://doi.org/10.1016/j.compenvurbsys.2023.101990
Received 19 October 2022; Received in revised form 12 May 2023; Accepted 15 May 2023
Available online 25 May 2023

intensity indicator. For instance, the scale indicators proposed by Olteanu et al. (2019) and Lan et al. (2020) are derived from the spatial entropy index, which would also be used for measuring segregation intensity (Theil & Finizza, 1971). When comparing cities, the variation observed in the different indicators is thus partly governed by the index they are based upon, which inhibits the robustness of a cross-sectional study. Measuring directly intensity, separation, and scale would allow to better disentangle these dimensions and to perform a meaningful comparison of segregation patterns across cities. This would unlock a great, yet untapped potential to study the different faces of segregation in cities, and to explore how certain segregation patterns associate with city characteristics.

While several studies propose metrics for intensity, separation, and scale in the literature; a method for measuring these dimensions separately is lacking. We address this methodological gap and then apply the proposed approach to all Dutch municipalities. Furthermore, we make two substantive contributions. First, we characterize segregation patterns in the Netherlands and explore similarities and differences across Dutch municipalities. Second, we demonstrate how the approach can be used to relate segregation patterns to several city characteristics, such as the city population, social mix, and income inequality, which have often been associated with segregation — see for example Moreno-Monroy and Veneri (2018); Gordon and Monastiriotis (2006); Krupka (2007); Natale, Scipioni, and Alessandrini (2018); Musterd, Marcińczak, Van, and Tammaru (2015).

In this study, we specifically investigate the spatial segregation of people with a non-western migration background in the Netherlands. Segregation between Dutch natives and the foreign-origin population is now a major issue for local and national authorities, as it is deemed to hamper integration and social mobility (Hartog & Zorlu, 2009; Tselios, Noback, van Dijk, & McCann, 2015; Vervoort, 2012; Zorlu & Mulder, 2008). Analyzing segregation from a multi-dimensional perspective is particularly suited in this context, as high segregation intensity associated with large scale is deemed to hamper integration (Tselios et al., 2015).

2. Method

To study segregation, this work considers the city fabric as an ensemble of segregated and mixed regions forming a certain segregation pattern. We propose to characterize and quantify residential segregation patterns by means of the following three metrics: intensity, separation, and scale. Intensity and separation relate to the spatial distribution of groups across regions, whereas scale relates to the size of the segregated regions. The proposed indicators are described and formally introduced in subsection 2.1. Before one can measure these indicators, it is necessary to first delineate segregated and mixed regions. To this end, we use a regionalization method to demarcate regions in which the potential to encounter individuals from each group of interest (potential exposure to each group) is uniform. We then identify a region as segregated if the potential exposure to a group in the region is significantly larger than average in the perimeter of study (subsection 2.2). The computation of the potential exposure is described in subsection 2.3. Fig. 1 below illustrates the different steps involved in the analysis, taking the municipality of Leiden as an example. The map on the left depicts the share of the group of interest in the population residing in each spatial unit (using raw demographic data). The middle map displays the exposure levels in each spatial unit, and the map on the right displays the regions where the potential exposure to the group of interest is significantly larger than the municipality average. Finally, intensity, separation, and scale can be used to summarize the characteristics of the segregated regions delineated on the right map in Fig. 1.

2.1. Indicators

This work proposes three indicators to characterize segregation of a group k in a city: intensity, separation, and scale. Together, they can express a wide diversity of segregation patterns. This subsection defines each metric formally and demonstrates the patterns they can capture using a simple example (Fig. 2). We assume here that segregated regions have already been identified and are available as input for subsequent analysis, as in the right map of Fig. 1 (see subsection 2.2 for more details). This study considers two groups, one of them being the group of interest. A region R is an ensemble of spatial units. It has one of the three following statuses, indicated by a variable s: the potential exposure to the group of interest is significantly lower than the city average ($s = -1$), higher than the city average ($s = 1$), or mixed ($s = 0$). The potential exposure to group k estimates the share of individuals from group k in the total population that one would encounter in a given space (spatial unit or region), in percentage terms.

2.1.1. Intensity

The segregation intensity of group k is the difference between the potential exposure to group k in all segregated regions $r \in U_{s=1}$ and the city average μ_k, expressed in percentage points (eq. 1). In this equation, n_R is the population of region R, $n_{U_{s=1}}$ is the population living in the ensemble of regions $U_{s=1}$, and π_{Rk} is the potential exposure to group k in region R. The intensity is therefore the average exposure in regions that

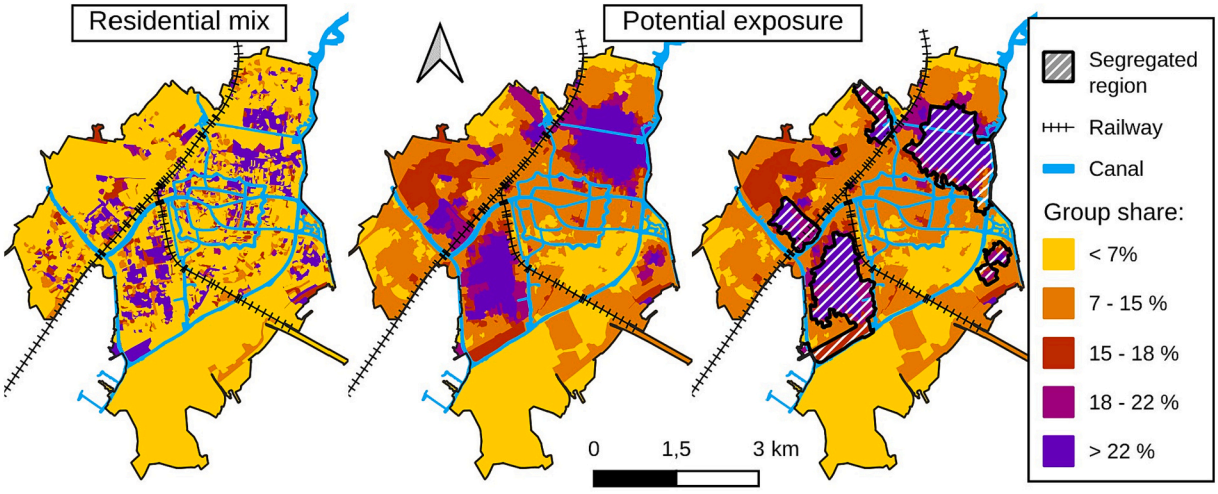


Fig. 1. Delineation of segregated regions in Leiden. The color indicates the proportion of individuals from the group of interest among the population either residing in the spatial unit (left map), or able to reach it (center and right map). The right map displays the regions identified as segregated.

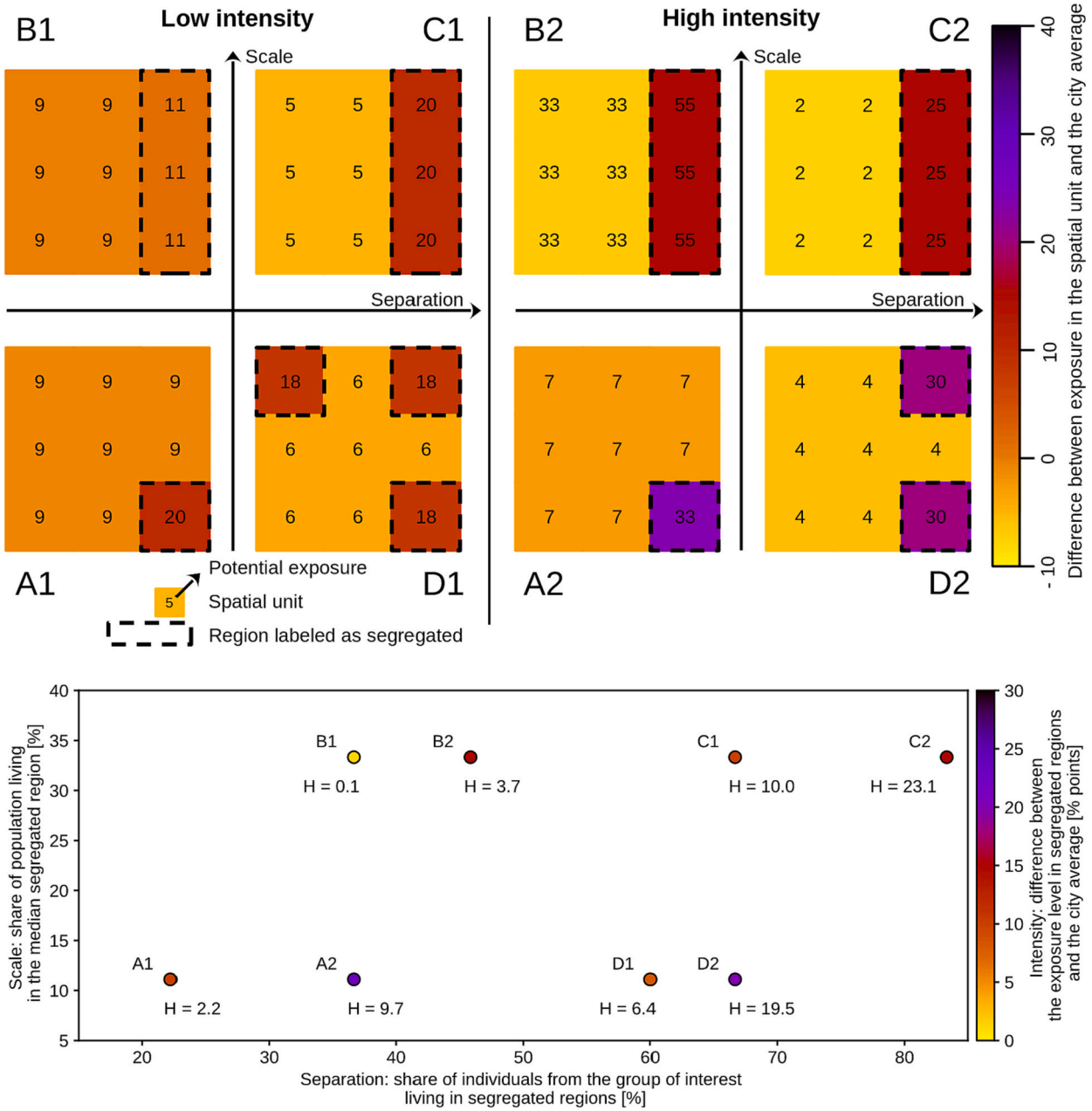


Fig. 2. Eight conceptual examples of cities with nine equally populated spatial units (top) demonstrating the diversity of segregation patterns that the intensity, separation, and scale indicators can capture in a three dimensional plot (bottom). In the conceptual examples, the value in a spatial unit indicates the potential exposure to the group of interest, while the color represents the difference between the potential exposure and the city average. Finally, regions overexposed to the group of interest are delineated by the dashed black borders. In the plot, the color represents the difference between the average potential exposure in these regions and the city average. In all settings except for setting B2, the average exposure in the city is set to 10% (setting B2 requires a higher city average in order to manifest itself). In the plot, the H value shown below each dot is the spatial entropy index corresponding to each segregation pattern.

are overexposed to group k, weighted by their population, compared to the city average. More details on how to obtain π_{Rk} — which reflects the extent to which region R is exposed to group k — are provided in subsection 2.3.

$$\phi_k = \sum_{R \in U_{s=1}} \frac{n_R}{n_{U_{s=1}}} \pi_{Rk} - \mu_k \tag{1}$$

2.1.2. Separation

The separation of group k is the proportion of individuals from this group living in a region that is overexposed to group k (eq. 2). It indicates the degree to which individuals from the group of interest live in a segregated region.

$$\chi_k = \frac{n_{k,U_{s=1}}}{n_{k,tot}} \tag{2}$$

2.1.3. Scale

The scale is the median size of segregated regions, in terms of population size. Half of the people living in a region where group k is overrepresented reside in a region that is more (or equally) populated than the scale indicator. It can be expressed in absolute terms (eq. 3), in number of inhabitants, or relative to the city population (eq. 4).

$$\psi_k = n_{med\{U_{s=1}\}} \tag{3}$$

$$\widetilde{\psi}_k = \frac{\psi_k}{n_{tot}} \tag{4}$$

2.1.4. Characterizing segregation patterns

The combination of these three distinctive indicators can cover a wide variety of residential segregation patterns. Fig. 2 displays eight hypothetical cities exhibiting different segregation patterns, and shows the intensity, separation, and scale for each setting in a three-dimensional plot. In addition, we also display for each setting the corresponding spatial entropy index *H*, a widely used 1-dimensional indicator for residential segregation (Theil & Finizza, 1971). The example is crafted so as to showcase two levels for each indicator (high and low) and demonstrate that all combinations between the indicators are in principle possible. For instance, situation *A2* represents a segregation pattern with high intensity, low scale, and low separation.

These two figures highlight the advantages of using three indicators over the *H* index alone. In the example, the *H* index can clearly differentiate segregation patterns where both separation and intensity are low (situations *A1* and *B1*), from patterns where they are both high (situations *C2* and *D2*): the *H* index is lower than 3% in the first case and higher than 15% in the second case. However, it does not allow to differentiate situations where separation is high and intensity is low (e.g. *C1*), from situations where separation is low and intensity is high (e.g. *A2*): the *H* index is the same for situations *C1* and *A2*, even though the two segregation patterns are remarkably different. Moreover, the *H* index value does not provide any information on the scale. Finally, it does not provide an intuitive representation of the segregation pattern. An *H* index of 4% for a city is hard to interpret for a policymaker if it is not benchmarked against other cities. The metrics proposed in this work explicitly target these drawbacks. They differentiate all eight proposed cases and can be intuitively communicated to policymakers. For instance, a separation index of 50% means that half of the people from the group of interest live in a region considered as overexposed to that group. Moreover, the different metrics are complementary, providing information on different aspects of urban segregation thereby making them suited for a meaningful cross-city comparison.

2.2. Delineation of spatially segregated regions

The indicators proposed in subsection 2.1 require a meaningful division of the urban fabric into segregated and mixed regions as input. This work uses a regionalization method, namely agglomerative clustering to aggregate contiguous spatial units into regions that are homogeneous in terms of potential exposure to the group of interest (subsection 2.2.1). Subsequently, regions in which the potential exposure is significantly larger than the city average are labeled as segregated (subsection 2.2.2).

2.2.1. Agglomerative clustering

We use agglomerative clustering to group spatial units into larger regions (Theodoridis & Koutroumbas, 2008). In the initialization phase, all spatial units are considered as independent clusters. Then, one merges clusters (hereafter called regions) iteratively. Only adjacent regions may be merged. For each iteration, one considers the dissimilarity between each pair of regions (in terms of potential exposure), and merges the two most similar regions. The similarity between regions is determined by the Ward distance, shown in eq. 5 measuring the increase in the within-region variance when two of them are merged (Müllner, 2011). In eq. 5, ν_i is the number of spatial units in region *i*.

$$d(i \cup j, k) = \sqrt{\frac{(\nu_i + \nu_k)\cdot d(i,k)^2 + (\nu_j + \nu_k)\cdot d(j,k)^2 - \nu_k \cdot d(i,j)^2}{\nu_i + \nu_j + \nu_k}} \tag{5}$$

Fig. 3 provides an example of agglomerative clustering. At first, all spatial units (*A, B, C, D,* and *E* on the left side of the figure) are treated as individual regions. The dissimilarity $d(i,j)$ between each pair (i,j) of these initial regions is the difference between the exposure level in the two spatial units. The most similar regions are *A* and *E*, but they cannot be merged because they are not adjacent. Instead, *A* and *B* are merged. Then the distance between the newly formed region $A \cup B$ and any other region *k* is computed using formula 5. We repeat the procedure until a stopping criterion is satisfied. The dendrogram in the right of Fig. 3 summarizes the successive merges, along with the corresponding increase in within-region variance.

When aggregating spatial units together, one should ensure that spatial units are adjacent, so that regions are spatially continuous. For that, we use a connectivity matrix to ban merging operations that would result in spatially discontinuous regions. The connectivity matrix Ω for a city with ν initial spatial units is a $\nu \times \nu$ matrix, in which component ω_{ij} is one if *i* is adjacent to *j*, and zero otherwise (see center of Fig. 3).

The agglomerative process is stopped when the increase in within-region variance exceeds a certain threshold. Like in any clustering procedure, one needs to find a suitable threshold that will strike a good balance between identifying many small and homogeneous regions or too few large and heterogeneous regions. If the threshold is set too high, the algorithm aggregates regions that do not have comparable exposure levels. If the threshold is set too low, the algorithm misses aggregating regions that should belong to the same larger region. We tune this parameter empirically by testing a wide range of values, and investigating the consistency of detected regions.

2.2.2. Labeling a region as segregated

The average exposure in region *R* is computed using eq. 6. In this equation, the potential exposure in spatial unit *i* is weighted by the population n_i residing in this spatial unit (variable θ_i represents the weight of spatial unit *i* in the region). We label region *R* as segregated if the average exposure in region $\overline{y_R}$ is significantly larger than the average exposure μ if the groups were to be randomly distributed in space. This is defined by referring to a difference of more than one standard deviation $\sigma_{\overline{y_R}}$, away from the average exposure in the city μ (see eq. 7). In this equation, the region *R* is segregated when *s* is 1 (overexposure) or -1 (underexposure). When *s* is 0, the region is considered as mixed. For details on how to calculate the standard deviation of exposure in a region, see appendix A.

$$\overline{y_R} = \frac{\sum_{i \in R} n_i y_i}{\sum_{i' \in R} n_{i'}} = \sum_{i \in R} \theta_i y_i \tag{6}$$

$$s = \begin{cases} -1 & \text{if } \dfrac{\overline{y_R} - \mu}{\sigma_{\overline{y_R}}} \leq -1 \\ 0 & \text{if } -1 < \dfrac{\overline{y_R} - \mu}{\sigma_{\overline{y_R}}} < 1 \\ 1 & \text{if } \dfrac{\overline{y_R} - \mu}{\sigma_{\overline{y_R}}} \geq 1 \end{cases} \tag{7}$$

Adjacent regions having the same value for *s* are merged.

2.3. Potential exposure

We compute the potential exposure in each spatial unit before aggregating them into homogeneous regions. The potential exposure to a given group in a spatial unit aims at representing the potential to encounter the group in the spatial unit, relative to the potential to encounter any group. One could also aggregate units using the residential mix (left map in Fig. 1). However, the residential mix is arguably less representative of the segregation experienced by individuals: if one spatial unit highly populated by a group is surrounded by units deserted by this group then the segregation experienced is most likely lower than what the residential mix would have indicated. Furthermore, the potential exposure also has a smoother distribution (center map in Fig. 1), which is more suited for aggregating units into regions (Spierenburg, van Cranenburgh, & Cats, 2022).

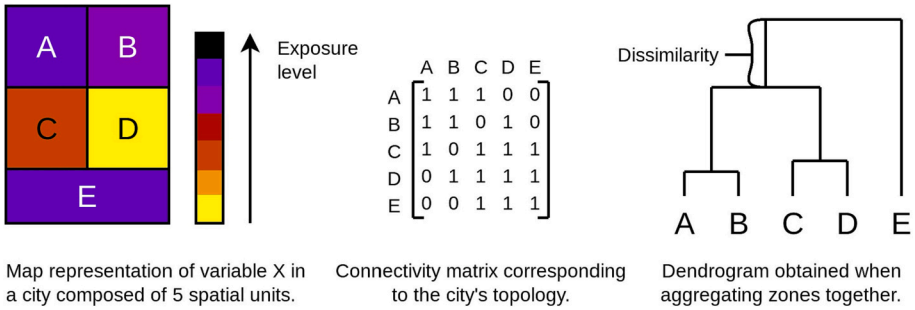


Fig. 3. Example of agglomerative clustering applied to a illustrative city (left). The connectivity matrix (center) indicates the spatial units that are adjacent. The dendrogram (right) depicts the successive merging operations between regions.

We use an accessibility metric to quantify the ease with which people from each group can reach the centroid of any given spatial unit. People able to reach the spatial unit are weighted using the walking time between their place of residence and the spatial unit based on the fundamental principle of human mobility patterns proposed by Schläpfer et al. (2021): the further away people live, the less likely they are to visit the spatial unit (subsection 2.3.1). Then, the potential exposure to a certain group in a spatial unit corresponds to the share of that group in the population able to reach this spatial unit (subsection 2.3.2).

2.3.1. Travel impedance

For a given destination spatial unit, we first determine the origin spatial units located within an acceptable walking distance from it. The shortest walking distance between any spatial unit and any other spatial unit is computed using the street layout. The walking time is calculated from the walking distance, using a walking speed of 4.5 km/h. Then, inhabitants living in the origin spatial units are weighted based on the walking time and the travel impedance function described by eq. 8.

$$w(t) = \begin{cases} 1 & \text{if } \ 0 \leq t[s] < 60 \\ \dfrac{3600}{t^2} & \text{if } \ 60 \leq t < 1200 \\ 0 & \text{if } \ t \geq 1200 \end{cases} \tag{8}$$

The travel impedance shown in eq. 8 is derived from the work of Schläpfer et al. (2021) providing a law to model the visitation pattern of individuals. This law is expressed in eq. 9. $\rho_i(t,f)$ is the influx of visitors coming to place i, living at travel time t[1] from i and visiting i at frequency f. α_i is a constant that depends on the destination place i, it relates to the attractiveness of i. η is set to 2 (derived empirically). The number of instances in which someone living at walking time t visits i at frequency f is $\rho_i(t,f) \cdot f$. Next, the number of instances in which someone living at a walking time t visits i is obtained from eqs. 10 and 11. We use this relation to model the number of visits of inhabitants from one spatial unit to another as a function of the walking time separating the two units (eq. 11). We set a cut-off of 20-min travel time to limit the number of shortest paths between spatial units to be computed, i.e. we stop exploring a path if the duration exceeds 20 min. All travel times shorter than 1 min are set to 1 min. The constant c_i does not affect the potential exposure indicator (see subsection 2.3.2). under the assumption that the attractiveness is constant across groups.

$$\rho_i(t,f) = \frac{\alpha_i}{(tf)^\eta} \tag{9}$$

$$w_i(t) = \int \rho_i(t,f) f df = \frac{\alpha_i}{t^2} \int \frac{1}{f} df \tag{10}$$

$$w_i(t) = \frac{c_i}{t^2} \quad \text{where} \quad c_i = \alpha_i \int \frac{1}{f} df \tag{11}$$

2.3.2. Potential exposure indicator

The likelihood that people from group k living in j visit spatial unit i is denoted by $w(t_{ij})$. The likelihood that people from group k visit destination i, noted q_{ik} is therefore the sum of likelihood over all origin spatial units (eq. 12). Then, π_{ik} the potential exposure to group k in spatial unit i is the share of visits from group k to i in all visits to spatial unit i. It can be observed that constant c_i does not affect the potential exposure indicator since it appears in both the numerator and the denominator of eq. 13.

$$q_{ik} = \sum_j \frac{c_i \cdot n_{jk}}{t_{ij}^2} = c_i \cdot \sum_j n_{jk} \Big/ t_{ij}^2 \tag{12}$$

$$\pi_{ik} = q_{ik} \Big/ \sum_{k'} q_{ik'} = \frac{c_i \cdot \sum_j n_{jk} \Big/ t_{ij}^2}{c_i \cdot \sum_{k'} \sum_j n_{jk'} \Big/ t_{ij}^2} \tag{13}$$

3. Case study and datasets

3.1. Case study

The method proposed in section 2 can be used to investigate segregation along any dimension such as income, age, or migration background; within a perimeter of study such as a city or a functional urban area (Dijkstra, Poelman, & Veneri, 2019). In the Netherlands, segregation of residents with a migration background persists and hampers their integration to the rest of the population (Hartog & Zorlu, 2009; Tselios et al., 2015; Vervoort, 2012; Zorlu & Mulder, 2008). In this paper, we therefore characterize residential segregation of the population with a non-western migration background in every Dutch municipality. The definition of the migration background used by the Centraal Bureau voor de Statistiek (van Leeuwen, 2020) is as follows. Inhabitants are grouped into three categories: individuals from Dutch descent (both parents were born in the Netherlands), individuals with a western migration background (Europe, North America, New Zealand, Australia, Japan, and Indonesia), and individuals with a non-western migration background. Individuals with a migration background originate from another country than the Netherlands or have one of their parents originating from another country than the Netherlands. We take Dutch municipalities, also called gemeenten, as perimeters of study, as they are local authorities responsible for urban planning and housing development of a well-demarcated area.

After characterizing residential segregation patterns for every Dutch municipality, we relate the observed patterns to three characteristics often associated with segregation: number of inhabitants, social mix, and income inequality (Gordon & Monastiriotis, 2006; Krupka, 2007; Moreno-Monroy & Veneri, 2018; Musterd et al., 2015; Natale et al.,

[1] In the work of Schläpfer et al. (2021), the law relates the influx $\rho_i(t,f)$ to the cartesian distance r rather than travel time t between two places.

2018).

3.2. Datasets

Demographic data are retrieved from the Centraal Bureau voor de Statistiek (van Leeuwen, 2020). The spatial units are the 6-digits postcodes. These units are around 100 × 100 m^2 large and populated by around 50 inhabitants in dense urban areas. This work uses data from 2017. The left map in Fig. 1 displays the population mix in each spatial unit.

Data on the street network are also required to compute walking times between spatial units. These are retrieved from the OpenStreetMap (2021) database, using the OSMnx library in Python to extract the street network in every municipality (Boeing, 2017).

In this work, we relate segregation patterns to the cities' population, social mix, and income inequality. The city population and the city's social mix are retrieved from the demographic dataset. Data on income inequality in Dutch municipalities are retrieved from the (Centraal Bureau voor de Statistiek (2019). The population in Dutch municipalities ranges from 1000 to 900,000 inhabitants, and around 80 of them are populated by more than 50,000 inhabitants. The share of individuals with a non-Western migration background ranges from 1 to 38% of the municipality population, with a median of 5%. Finally, Dutch municipalities show relatively low income inequality compared to the rest of OECD countries (Veneri et al., 2021).

3.3. Reflection on the modifiable areal unit problem

The demographic data used in this analysis are provided for spatial units defined by the data supplier. Regionalization processes are generally subject to the modifiable areal unit problem (MAUP) when using such data (Openshaw, 1981). Fig. 4 illustrates the impact of the MAUP on the outcome of a regionalization process. This figure depicts a city inhabited by two household groups, A (purple) and B (yellow). The data supplier provides demographic data using a certain partition of space into spatial units, such as grid 1 or grid 2 (top and bottom left maps in Fig. 4). In these two examples, each spatial unit (black squares) contains four households. The middle top and bottom maps of Fig. 4 represent the spatial demographic datasets resulting from the two proposed partitions of space, where the color of a spatial unit indicates the share of households from group A residing in it. The data supplier would provide such a dataset. In these two maps, the hatched areas represent regions identified as segregated when clustering spatial units on demographic data. In this case, the outcome of the regionalization method is sensitive to the grid used to partition space: regions obtained using data from grid 1 are different from the ones obtained using data from grid 2.

Our method partially mitigates this issue, as we cluster spatial units on potential exposure instead of demographic data. The top and bottom right map in Fig. 4 represent the potential exposure to group A in each spatial unit, using data from grid 1 and data from grid 2. The potential exposure can be seen as a spatial moving average of demographic data, which filters out small-scale differences between data from grid 1 and data from grid 2 (Spierenburg et al., 2022). Consequently, regions obtained by clustering spatial units on potential exposure are less sensitive to the MAUP than the ones obtained by clustering units based on raw demographic data. In Fig. 4, regions identified as segregated are more consistent on the right-hand maps than on the middle maps.

The size of spatial units also affects the outcome of a regionalization process. More specifically, the minimum size of regions delineated by our method is constrained by the spatial resolution of the demographic dataset (100 × 100 m^2 in our case study). A better resolution could be obtained using the micro-data from the Dutch statistics institute, providing data at the household level (Centraal Bureau voor de Statistiek, 2021). However, such data is not openly available. Our use of the postcode data enables the reproducibility of the results.

3.4. Code and data produced

The computational workflow supporting this publication is executed via a set of 11 python scripts. We have published the code (MIT license), along with the data produced at each stage of the computational workflow (CC-BY-4.0 license) on the following platform: https://doi.org/10.4121/c.6237276.

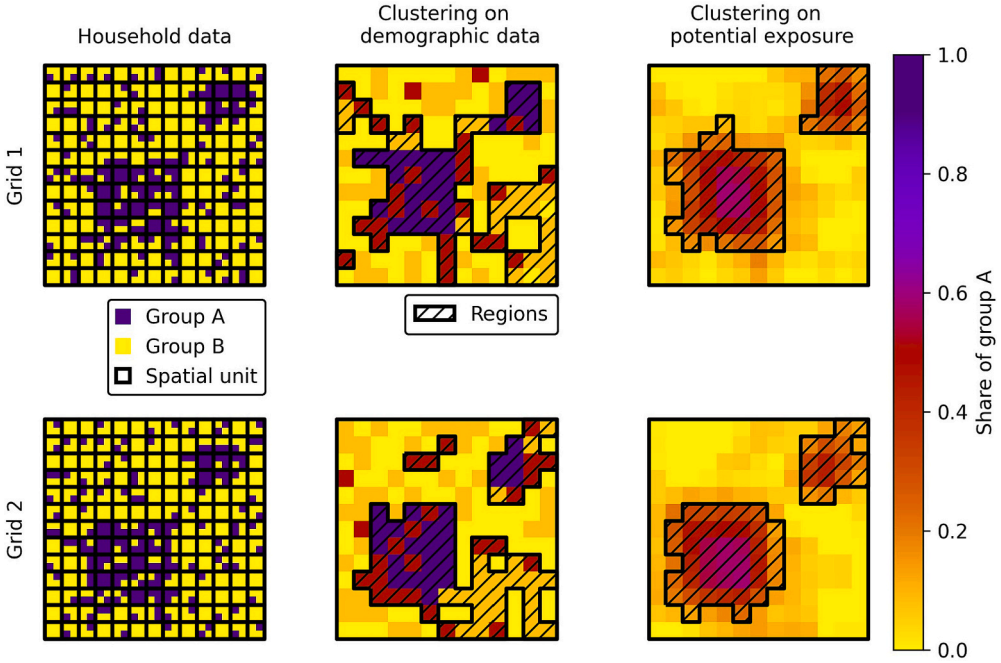


Fig. 4. Clustering spatial units into regions based on demographic data (middle) and on potential exposure (right), using two different definitions of spatial units (grid 1 and grid 2 on the left).

4. Results

The method described in section 2 proposes three components of residential segregation: intensity, separation, and scale. The combination of these components allows a multi-faced identification of similarities and differences in the segregation patterns that prevail across municipalities. In this section, we showcase this potential by exploring the variety of segregation patterns in all Dutch municipalities[2] (subsection 4.1), and relating the observed patterns to municipality characteristics often associated with segregation in the literature (subsection 4.2) (Gordon & Monastiriotis, 2006; Krupka, 2007; Moreno-Monroy & Veneri, 2018; Musterd et al., 2015; Natale et al., 2018).

4.1. Characterization of segregation patterns in the Netherlands

Residential segregation can potentially exhibit a wide range of spatial patterns, as suggested in the hypothetical example provided in Fig. 2, motivating the need to study segregation as a multidimensional phenomenon. In the following, we employ the three proposed measures of segregation to examine how those segregation patterns manifest themselves across Dutch municipalities. This enables determining which ones of the conceptual examples from Fig. 2 are actually encountered in practice, and identify related trends among Dutch municipalities. Fig. 5 plots the distributions of every indicator across Dutch municipalities, whereas Fig. 6 shows how they relate to one another. Fig. 7 summarizes the three indicators for all municipalities in one plot, reproducing Fig. 2 that was used in the simple example provided in subsection 2.1.4.

4.1.1. Descriptive statistics

Fig. 5, depicts the distributions of each segregation indicator across Dutch municipalities. Most Dutch municipalities (75%) show low intensity (lower than 10 percentage points above the municipality exposure average), and yet a high level of separation (more than 50% of the group of interest live in segregated regions). Dutch municipalities do not show particular trends regarding the scale of segregation relative to their size. Small, medium, and large-scale segregation occur in relatively even proportions.

The shapes of the distributions of intensity, separation, and relative scale show high differences when considering skewness and kurtosis. Intensity is right-skewed with a long tail: the skewness is positive (1.4), and the excess kurtosis is large (2.2). The relative scale is also right-skewed (skew of 0.6) with low kurtosis (−0.1). Separation is in contrast left-skewed (skew of −1.2) with low excess kurtosis (0.7). Summarizing residential segregation with a uni-dimensional indicator conceals the disparity in the distributions of the three individual dimensions, and this result already motivates the need to study this phenomenon as multi-dimensional. The maps in Fig. 5 do not directly display any clear geographic trend in relation to the three indicators.

Fig. 6 displays how the three indicators relate to one another. Relative scale and separation are strongly and positively correlated ($\rho = 0.71$), whereas intensity exercises no to limited negative Pearson correlation with separation ($\rho = -0.09$) and relative scale ($\rho = -0.22$), respectively.[3] The spatial entropy index combines intensity and separation, it is low (respectively high) when intensity and separation are low (respectively high). This indicator confounds situations where both indicators have diverging values (high separation low intensity, and low separation high intensity), as shown in subsection 2.1.4. Yet, our results show that such situation happens in practice. High separation does not necessarily come with high intensity, and conversely (no positive Pearson correlation observed), questioning the robustness of such indicator combining these two dimensions.

4.1.2. The diversity of segregation patterns in Dutch municipalities

Dutch municipalities exhibit a rich diversity of segregation patterns. Fig. 7 displays the three indicators together in a single plot, along with three examples of municipalities characterized by substantially different segregation patterns. Some municipalities like Losser at the bottom left of the figure show low intensity, low separation, and low relative scale, some at the top right like the Hague show high intensity, high separation, and high relative scale, while some others like Alkmaar show high intensity, high separation, and low relative scale. These examples illustrate that all three dimensions add information when investigating segregation in the Netherlands.

Importantly, this result highlights the deficiencies of a uni-dimensional indicator like the spatial entropy index which masks differences in features of segregation that remain uncovered. The spatial entropy index does not capture the scale, and collapses intensity and separation indicators into a single value. A segregation pattern expressing high intensity and low separation can have the same value for the spatial entropy index as another pattern with low separation and high intensity (see Fig. 2 in subsection 2.1.4), while our empirical results in Figs. 6 and 7 demonstrate that both patterns exist in reality, and that intensity and separation are uncorrelated ($\rho = -0.09$). This underscores the need to consider intensity and separation independently.

Fig. 7 conveys another finding about residential segregation in the Netherlands. Municipalities do not converge towards any particular pattern. Six of the eight conceptual examples illustrated in Fig. 2, are observed in the Netherlands, and each of which prevails in a number of municipalities. Dutch municipalities only share one common characteristic: they do not exhibit the combination of a low separation level and large relative scale level (situations *B1* and *B2* in Fig. 2), hence there is no observation in the top left quadrant of Fig. 7. Notwithstanding, Dutch municipalities show little similarities and express high variability across all dimensions considered. Therefore, we investigate in the subsequent subsection 4.2 whether, despite the variety of segregation patterns observed, municipalities sharing common characteristics also manifest similar patterns. For instance, Fig. 7 suggests that the most populated municipalities are characterized by a higher separation level (none of the large municipalities appears in the left half of the plot).

4.2. Association with municipality characteristics

Subsection 4.1 showed that residential segregation takes various configurations in the Netherlands, and that no general segregation pattern prevails in Dutch municipalities. However, these municipalities also differ in many other aspects such as their population size, demography, or economy. We next turn into examining the association between the observed segregation patterns and municipality characteristics. In particular, we investigate this by relating intensity, separation, and relative scale to three characteristics often associated with segregation in the literature: municipality size, share of the group of interest in the municipality population (hereafter called social mix), and income inequality (Gordon & Monastiriotis, 2006; Krupka, 2007; Moreno-Monroy & Veneri, 2018; Natale et al., 2018). In this work, municipality size is measured using the total population living in the municipality, the social mix is the proportion of residents with a non-Western migration background living within the municipal borders, and income inequality is represented by the local Gini coefficient. Fig. 8 plots the three segregation indicators against the three municipal characteristics considered.

[2] Municipalities defined as of 2017, excluding Baarle-Nassau, a municipality with erratic borders between the Netherlands and Belgium.

[3] We have removed 9 municipalities with extremely high intensity (greater than 0.3) when computing these correlation coefficients: Gulpen Wittem, Eijsden Margraten, Laarbeek, Schinnen, Tynaarlo, Binnenmaas, Noordenveld, Vlagtwedde, het Bildt. In these municipalities, regions labeled as segregated are asylum shelters and prisons. The social mix observed in these regions is not due to segregation. These outliers show extreme values for the intensity (more than 40%), as the share of individuals with a migration background is lower than 2%, in these municipalities.

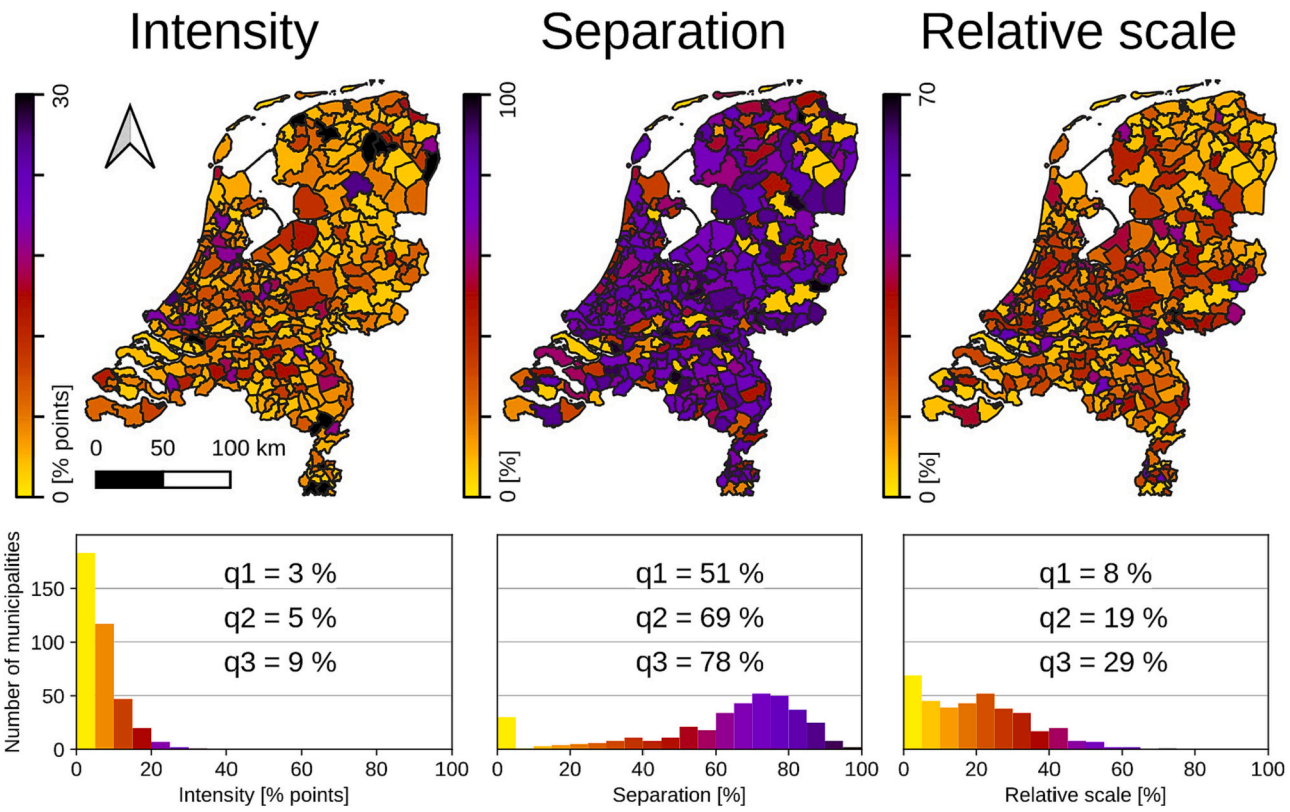


Fig. 5. Distribution of the intensity, separation, and relative scale indexes. The color represents which interquartile the observations belong to for a given indicator.

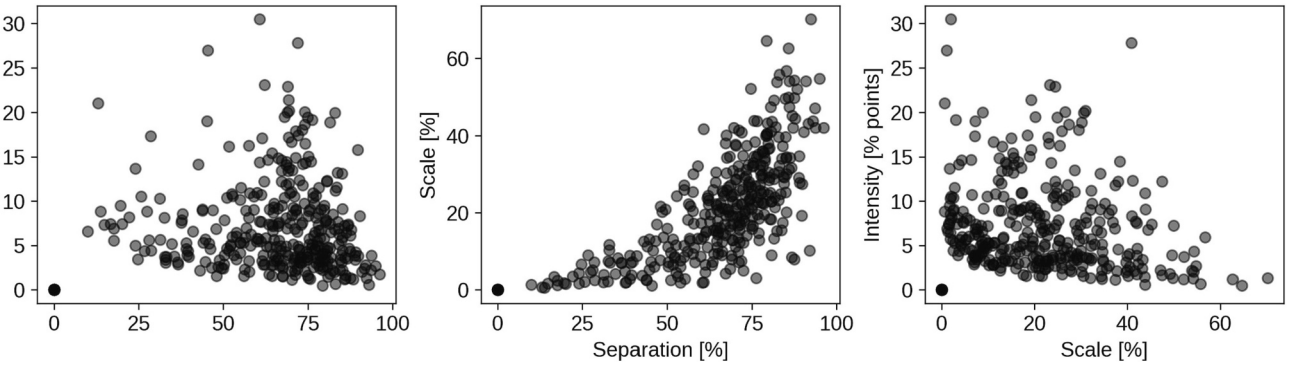


Fig. 6. Relation between the three different segregation indicators.

Larger, more mixed, and more unequal municipalities show higher levels of segregation intensity and separation. Intensity is positively correlated with all municipal characteristics: population ($\rho = 0.42$), social mix ($\rho = 0.64$), and the Gini coefficient ($\rho = 0.16$). Even though separation shows lower Pearson correlation values with these three characteristics (0.15, 0.30, and 0.01, respectively), a clear elbow-shape relation can be observed in Fig. 8. However, the association of intensity and separation with municipality population, social mix, and income inequality seem to be the manifestations of a similar underlying phenomenon. First, these characteristics are themselves correlated. Larger municipalities tend in the Netherlands (as well as elsewhere, see Natale et al. (2018) and Castells-Quintana, Royuela, and Veneri (2020)) to be more mixed and more unequal. Second, the relations between intensity (respectively separation) and municipality population, social mix, and income inequality follow the same linear (respectively elbow-shape) trend in Fig. 8. Hence, researchers exploring the relation between one of these characteristics and segregation should be particularly prudent and control for a potential omitted variable bias.

We find that the scale of segregation relative to the municipality population does not correlate with any of the municipal characteristics considered in Fig. 8. Conversely, the absolute scale, represented by the number of inhabitants living in the median segregated region (see subsection 2.1.3), is remarkably linearly associated with the municipality population (right plot in Fig. 9). The Pearson correlation coefficient is 0.93. One could expect some correlation, as the indicator for the absolute scale has an upper bound that is constrained by the municipality population: it cannot be larger than the total population. Such dependence would usually result in a correlation. However, the relation does not necessarily have to be linear, it could have been sub-linear, if larger municipalities were to be characterized by more segregated regions, rather than larger regions. Such a situation would have resulted in a concave shape of the right plot in Fig. 9. The fact that the right plot in Fig. 9 shows a clear linear relationship between the absolute scale of segregation and the municipality size, suggests that large municipalities

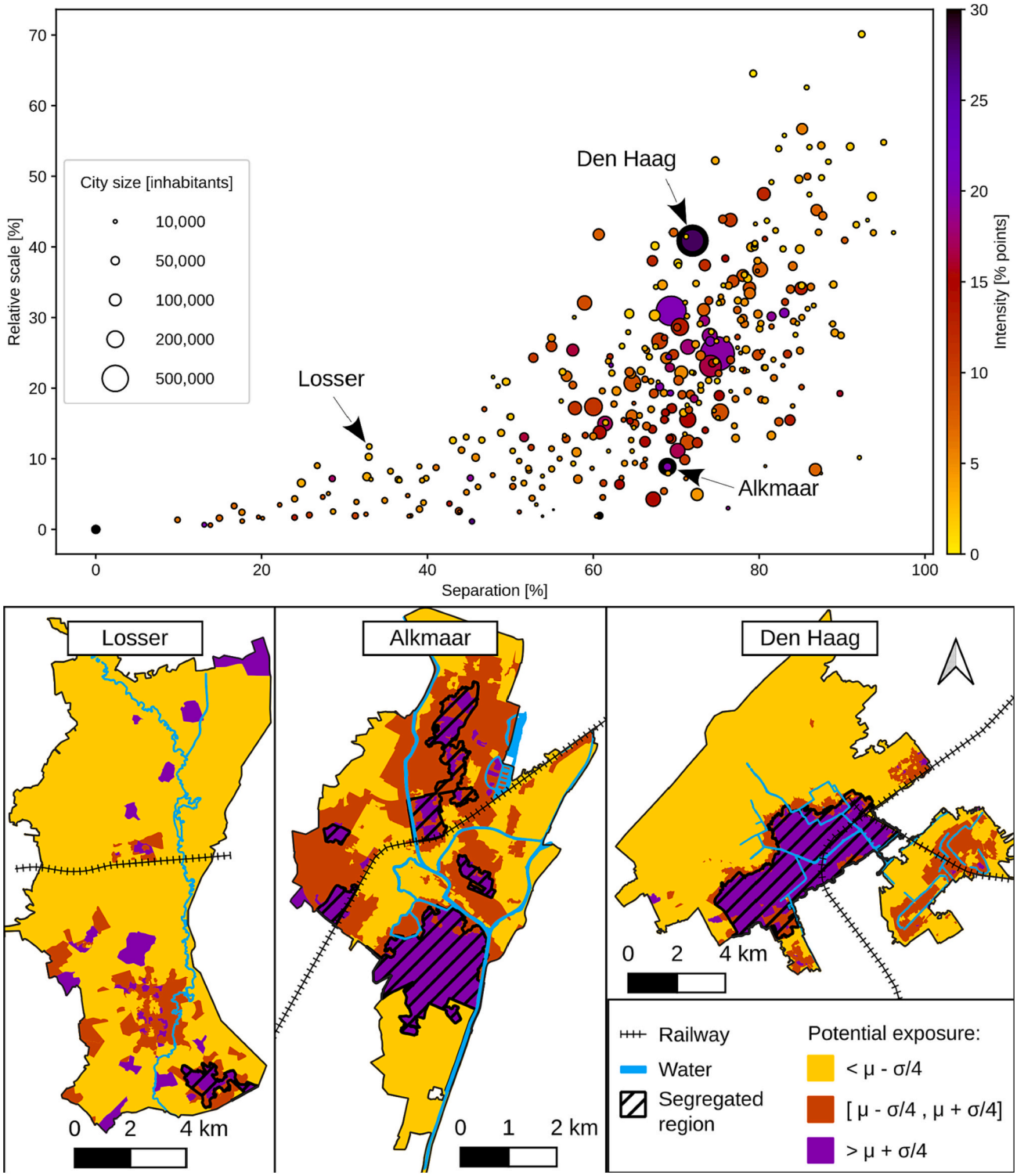


Fig. 7. Relation between segregation indicators in Dutch municipalities (top plot), along with three observed segregation patterns (bottom three maps). The separation and the relative scale are shown on the x and y axes, respectively, while the color of a point depicts the segregation intensity. The size of points is proportional to the municipality population. The three maps display the exposure levels per spatial unit, along with the identified segregated regions in the municipalities of Losser, Alkmaar, and Den Haag. The color indicates the value of the exposure in a spatial unit compared to the municipality average μ based on the standard deviation σ. An interactive version of this figure can be found at https://lucas-spierenburg.eu/projects/charact_seg_patterns/seg_pat_nth.html.

are (proportionally) larger versions of small municipalities, as far as the scale of segregation is considered.

5. Conclusion

In this paper, we propose a method to characterize residential segregation along three dimensions: intensity, separation, and scale. This method quantifies each dimension separately, which allows to

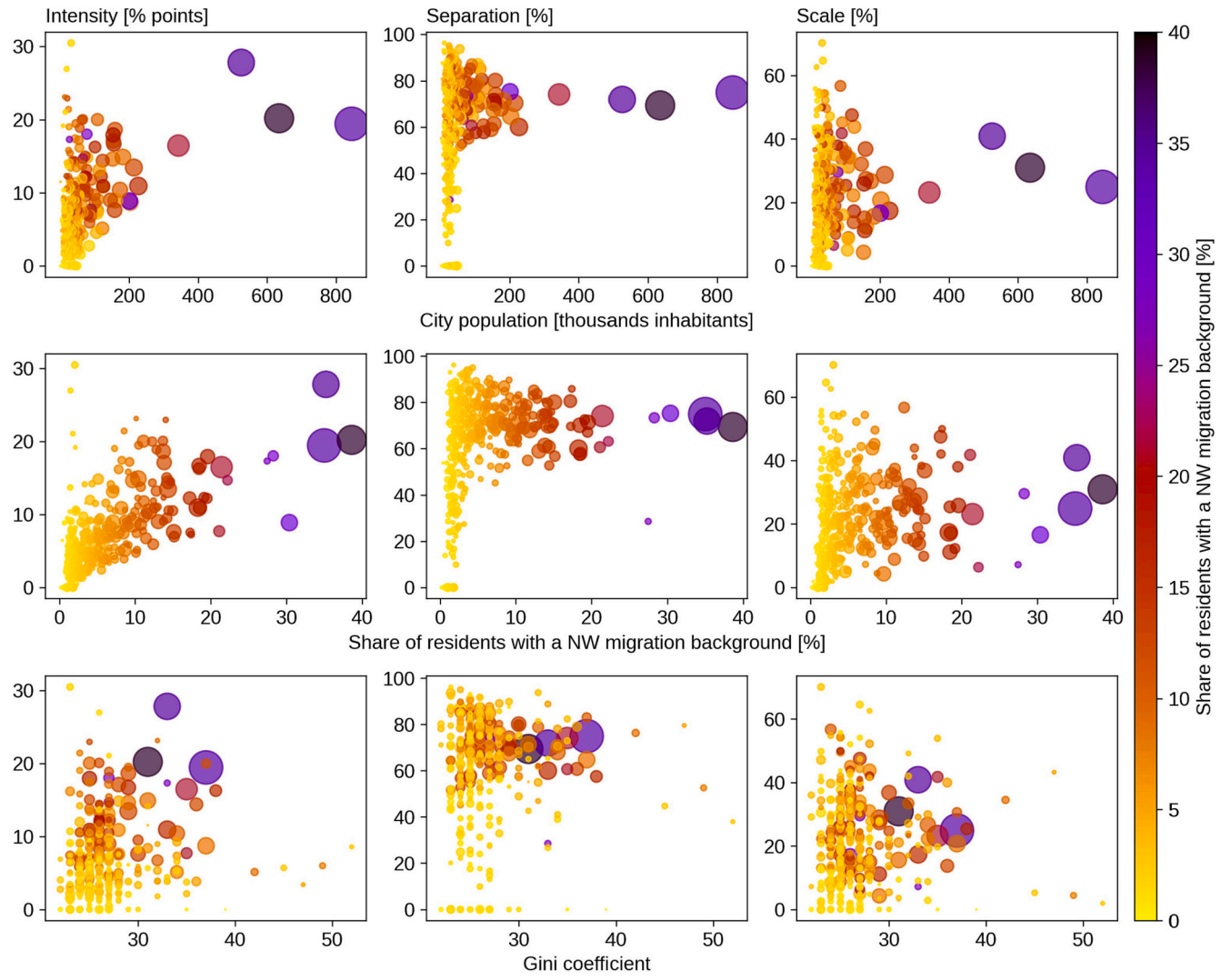


Fig. 8. Relation between intensity (left column), separation (middle column), relative scale (right column) indicators and municipality size (top row), social mix (middle row), and Gini coefficient (bottom row). The color of an observation represents the social mix. The size of each observation is proportional to the municipality population. An interactive version of this figure can be found at https://lucas-spierenburg.eu/projects/charact_seg_patterns/assoc_city_char.html.

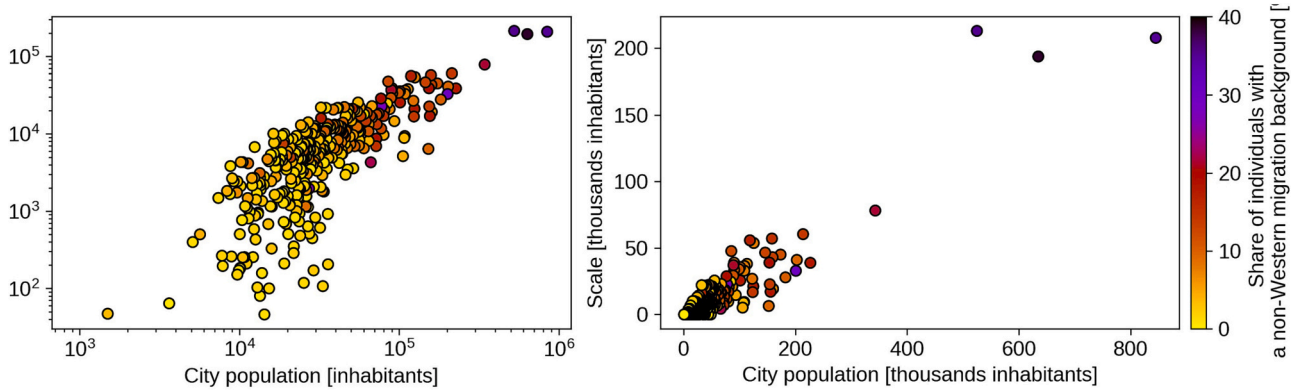


Fig. 9. Relation between the absolute scale of segregation (population in the median segregated region), and municipality population. The color of an observation represents the share of residents with a non-Western migration background. As the population in Dutch municipalities is such that there are many small municipalities and few large municipalities, we plot the relation in a linear-linear (right) and a log-log graph (left).

discern every combination of the three indicators. We demonstrate it in a Dutch case study where we could successfully identify a wide variety of residential segregation patterns, and relate these patterns to city characteristics.

We reveal two trends in residential segregation patterns in the Netherlands. First, we observe that both intensity and separation levels

associate with municipality size, income inequality, and migrant share. While this result is in line with past research (Gordon & Monastiriotis, 2006; Natale et al., 2018; Veneri et al., 2021), it also suggests that the association of segregation with city size, migrant share, and income inequality are three faces of the same phenomenon. These three variables are correlated in the Netherlands, and all of which are similarly and positively associated with segregation intensity (linear relation) and separation (concave relation). Second, the absolute scale of segregation is highly correlated with municipality size. This implies that larger municipalities have larger segregated regions rather than more segregated regions, when compared to smaller municipalities.

Our study provides two arguments to prefer a multi-dimensional over a uni-dimensional analysis of residential segregation. First, uni-dimensional metrics, such as the spatial entropy index, do not capture the spatial scale, while our result shows that scale varies substantially across Dutch municipalities, which is consistent with the results of Petrović et al. (2018). Second, such uni-dimensional metric usually combines intensity and separation, which are in practice negatively correlated in the Netherlands.

In addition to providing a more complete characterization of residential segregation, our indicators are easier to comprehend than the spatial entropy index, which cannot be interpreted directly (Theil & Finizza, 1971). Our indicators relate to actual quantities such as the proportion of individuals living in segregated regions for the separation index. This feature makes them more understandable not only for policymakers, but also for a broader audience and thereby may stimulate a more evidence-based public debate.

The full potential of the method is yet untapped. Further research could pursue the association of segregation patterns with geographic, demographic, and urban characteristics; investigate other social dimensions such as education or income; evaluate how segregation patterns evolve over time; or assess how they vary across countries. Exploring these research directions would enrich the understanding of segregation, and, hopefully, assist local and national authorities in designing more inclusive cities.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data and code

All raw data, processed data, and results, along with the source code can be accessed on the following link: https://doi.org/10.4121/c.6237276

CRediT authorship contribution statement

Lucas Spierenburg: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Sander van Cranenburgh:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Oded Cats:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

None.

Data availability

Data and code available at: https://doi.org/10.4121/c.6237276

Acknowledgment

This work is supported by the TU Delft AI initiative.

Appendix A. Computing the variance of the average exposure in a region

The variance of the average exposure in a region $\overline{y_R}$ can be computed analytically from the equations below.

$$\overline{y_R} = \sum_{i \in R} \theta_i y_i \tag{A.1}$$

$$Var(\overline{y_R}) = \sum_{i \in R} \sum_{j \in R} Cov(\theta_i y_i, \theta_j y_j) \tag{A.2}$$

$$Var(\overline{y_R}) = \sum_{i \in R} \sum_{j \in R} \theta_i \theta_j Cov(y_i, y_j) \tag{A.3}$$

The coefficients θ_i are computed from eq. A.4.

$$\theta_i = \frac{n_i}{\sum_{i' \in R} n_{i'}} \tag{A.4}$$

The covariance matrix Σ containing the covariances $Cov(y_i, y_j)$ can be derived analytically, from the definition of y_i (see eq. A.5). In this equation, $w(t_{ik})$ is the travel impedance when walking from k to zones i, and x_k is the share of individual from the group of interest residing in zone k. The derivation of the covariance coefficients can be found in eqs. A.7 to A.9.

$$y_i = \frac{\sum_k w(t_{ik}) \cdot n_k \cdot x_k}{\sum_{k'} w(t_{ik'}) \cdot n_i} \tag{A.5}$$

$$y_i = \sum_k c_{ik} x_k \tag{A.6}$$

$$Cov(y_i, y_j) = Cov\left(\left(\sum_k c_{ik} x_k \right), \left(\sum_{k'} c_{jk'} x_{k'} \right) \right) \tag{A.7}$$

$$Cov(y_i, y_j) = \sum_k \sum_{k'} c_{ik} \cdot c_{jk'} \cdot Cov(x_k, x_{k'}) \tag{A.8}$$

$$Cov(y_i, y_j) = \sum_k c_{ik} \cdot c_{jk} \cdot Var(x_k) \tag{A.9}$$

In eq. A.8, the covariance of the two variables x_k and $x_{k'}$ is given in eq. 6.10. To assess the significance of the average $\overline{y_R}$ in relation to μ in eq. 7, we compare to a case where the variables x_k are randomly allocated to the zone k (the population n_k and the travel impedance $w(t_{ik})$ remain the same). This allows to assess the extend to which the average $\overline{y_R}$ can be observed by luck. The variables x_k are randomly allocated, they are therefore independent. The covariance $Cov(x_k, x_{k'})$ is 0 when $k \neq k'$. The variance $Var(x_k)$ is obtained from the actual distribution of x (the share of individuals from the group of interest residing in a zone).

$$Cov(x_k, x_{k'}) = \begin{cases} 0 & \text{if } k \neq k' \\ Var(k) & \text{if } k = k' \end{cases} \quad \text{(Uncorrelated variables)} \tag{A.10}$$

The value of coefficient c_{ik} is expressed in eq. A.11.

$$c_{ik} = \frac{w(t_{ik}) \cdot n_k}{\sum_{k'} w(t_{ik'}) \cdot n_{k'}} \tag{A.11}$$

This computation can be summarized by the matrix multiplication shown in eq. A.12, where C is given in eq. A.13.

$$\Sigma = C^T \times C \cdot \sigma_x^2 \tag{A.12}$$

$$C = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1N} \\ c_{21} & & & \\ \vdots & & & \\ c_{N1} & & & c_{NN} \end{bmatrix} \tag{A.13}$$

References

Andersson, E. K., Malmberg, B., Costa, R., Sleutjes, B., Marcin Stonawski, J., & Valk, H. A. G. D. (2018). A comparative study of segregation patterns in Belgium, Denmark, the Netherlands and Sweden: Neighbourhood concentration and representation of non-European migrants. *European Journal of Population, 34*, 251–275.

Boeing, G. (2017). Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems, 65*, 126–139.

Castells-Quintana, D., Royuela, V., & Veneri, P. (2020). Inequality and city size: An analysis for oecd functional urban areas. *Papers in Regional Science, 99*(4), 1045–1064.

Centraal Bureau voor de Statistiek. (2019). 8.1.5 ongelijkheid (gini-coëfficiënt) inkomen, 2017. Retrieved from https://longreads.cbs.nl/welvaartinnederland-2019/ongelijkheid-in-inkomen-en-vermogen/. in July 2022.

Centraal Bureau voor de Statistiek. (2021). Microdata: Zelf onderzoek doen. https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-enmicrodata/microdata-zelf-onderzoek-doen. last accessed in June 2021.

Clark, W. A., Anderson, E., Östh, J., & Malmberg, B. (2015). A multiscalar analysis of neighborhood composition in los Angeles, 2000–2010: a location-based approach to segregation and diversity. *Annals of the Association of American Geographers, 105*, 1260–1284.

Dijkstra, L., Poelman, H., & Veneri, P. (2019). The EU-OECD definition of a functional urban area. Data available at https://www.oecd.org/cfe/regionaldevelopment/functional-urban-areas.htm accessed in June 2020.

Durlauf, S. N. (1996). A theory of persistent income inequality. *Journal of Economic Growth, 1*, 75–93.

Fowler, C. S. (2016). Segregation as a multiscalar phenomenon and its implications for neighborhood-scale research: The case of South Seattle 1990-2010. *Urban Geography, 37*, 1–25.

Gemeente Amsterdam. (2017). Woonagenda 2025. https://www.amsterdam.nl/en/policy/urban-development/living/.

Gordon, I., & Monastiriotis, V. (2006). Urban size, spatial segregation and inequality in educational outcomes. *Urban Studies, 43*, 213–236.

Gramberg, P. (1998). School segregation: The case of Amsterdam. *Urban Studies, 35*(3), 547–564.

Hartog, J., & Zorlu, A. (2009). Part 1: Ethnicity, religion, discrimination and exclusion: Ethnic segregation in the Netherlands: An analysis at neighbourhood level. *International Journal of Manpower, 30*, 15–25.

Krupka, D. J. (2007). Are big cities more segregated? Neighbourhood scale and the measurement of segregation. *Urban Studies, 44*, 187–197.

Lan, T., Kandt, J., & Longley, P. (2020). Geographic scales of residential segregation in English cities. *Urban Geography, 41*, 103–123.

van Leeuwen, N. (2020). *Statistische gegevens per vierkant en postcode 2019â€2018â€2017*. Centraal Bureau voor de Statistiek. Retrieved from https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische-data/gegevens-per-postcode. in June 2021.

Ludwig, J., Duncan, G. J., Gennetian, L. A., Katz, L. F., Kessler, R. C., Kling, J. R., & Sanbonmatsu, L. (2012). Neighborhood effects on the long-term well-being of low-income adults. *Science, 337*.

Massey, D. S., & Denton, N. A. (1988). The dimensions of residential segregation. *Social Forces, 67*(2), 281–315.

Mayor of London. (2021). The London Plan: the spatial development strategy for Greater London. https://www.london.gov.uk/what-we-do/planning/london-plan/new-london-plan/london-plan-2021.

Moreno-Monroy, A. I., & Veneri, P. (2018). *Divided cities: Understanding intra-urban inequalities, chapter introduction*. OECD Publishing.

Müllner, D. (2011). *Modern hierarchical, agglomerative clustering algorithms. arXiv*.

Musterd, S., Marcińczak, S., Van, M., & Tammaru, H. T. (2015). *Socio-economic segregation in European capital cities: Increasing separation between poor and rich. Technical report*. Institute of the Study of Labor, IZA.

Natale, F., Scipioni, M., & Alessandrini, A. (2018). *Divided cities: Understanding intra-urban inequalities, chapter spatial segregation of migrants in EU cities*. OECD Publishing.

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. *Urban Studies, 57*, 176–197.

Olteanu, M., Randon-Furling, J., & Clark, W. A. (2019). Segregation through the multiscalar lens. *Proceedings of the National Academy of Sciences, 116*(25), 12250–12254.

Openshaw, S. (1981). The modifiable areal unit problem. In *Quantitative geography: a British view* (pp. 60–69).

OpenStreetMap. (2021). OpenStreetMap data. Last accessed in June 2021, more info at https://www.openstreetmap.org/.

Östh, J., Clark, W. A., & Malmberg, B. (2015). Measuring the scale of segregation using k-nearest neighbor aggregates. *Geographical Analysis, 47*, 34–49.

Owens, A. (2018). Income segregation between school districts and inequality in students' achievement. *Sociology of Education, 91*, 1–27.

Petrović, A., van Ham, M., & Manley, D. (2018). Multiscale measures of population: Within- and between-city variation in exposure to the sociospatial context. *Annals of the American Association of Geographers, 108*(4), 1057–1074.

Reardon, S. F., & O'Sullivan, D. (2004). Measures of spatial segregation. *Sociological Methodology, 34*, 121–162.

Scarpa, S. (2015). The impact of income inequality on economic residential segregation: The case of Malmö, 1991–2010. *Urban Studies, 52*, 906–922.

Schläpfer, M., Dong, L., O'Keeffe, K., Santi, P., Szell, M., Salat, H., … West, G. B. (2021). The universal visitation law of human mobility. *Nature, 593*, 522–527.

Spierenburg, L., van Cranenburgh, S., & Cats, O. (2022). A regionalization method filtering out small-scale spatial fluctuations. In *, 3. AGILE: GIScience Series* (p. 61).

Tammaru, T., van Ham, M., Marcińczak, S., & Musterd, S. (2015). *Socio-economic segregation in European capital cities: East meets West*. Routledge.

Taylor, C., & Gorard, S. (2001). The role of residence in school segregation: Placing the impact of parental choice in perspective. *Environment and Planning A: Economy and Space, 33*(10), 1829–1852.

the United Nations. (2017). New Urban Agenda. https://habitat3.org/the-new-urban-agenda.

Theil, H., & Finizza, A. J. (1971). A note on the measurement of racial integration of schools by means of informational concepts. *The Journal of Mathematical Sociology, 1* (2), 187–193.

Theodoridis, S., & Koutroumbas, K. (2008). *Pattern recognition, chapter 13: Clustering algorithms II: Hierarchical algorithms* (4th ed., pp. 654–658). Academic Press.

Tóth, G., Wachs, J., Clemente, R. D., Jakobi, Á., Ságvári, B., Kertész, J., & Lengyel, B. (2021). Inequality is rising where social network segregation interacts with urban topology. *Nature Communications, 12*.

Tselios, V., Noback, I., van Dijk, J., & McCann, P. (2015). Integration of immigrants, bridging social capital, ethnicity, and locality. *Journal of Regional Science, 55*(3), 416–441.

Veneri, P., Comandon, A., Àngel Garcia-López, M., & Daams, M. N. (2021). What do divided cities have in common? An international comparison of income segregation. *Journal of Regional Science, 61*, 162–188.

Vervoort, M. (2012). Ethnic concentration in the neighbourhood and ethnic minorities' social integration: Weak and strong social ties examined. *Urban Studies, 49*, 897–915.

Ville de Paris. (2021). Charte pour un aménagement durable et inclusif de Paris. https://cdn.paris.fr/paris/2021/03/02/343e9528c3a88309b30920627c486c97.pdf.

Wang, Q., Phillips, N. E., Small, M. L., & Sampson, R. J. (2018). Urban mobility and neighborhood isolation in America's 50 largest cities. *Proceedings of the National Academy of Sciences of the United States of America, 115*, 7735–7740.

Williams, D. R., & Collins, C. (2001). Racial residential segregation:A fundamental cause of racial disparities in health. *Public Health Reports, 116*, 404–416.

Zorlu, A., & Mulder, C. H. (2008). Initial and subsequent location choices of immigrants to the Netherlands. *Regional Studies, 42*, 245–264.