Title
Measuring the scale of segregation using k-nearest neighbor Aggregates

Journal
Geographical Analysis, 47(1)

ISSN
0016-7363

Authors
O sth, J
Clark, WAV
Malmberg, B

Publication Date
2014

DOI
10.1111/gean.12053

Peer reviewed

geographical analysis

Geographical Analysis (2014), 1–16 Geographical Analysis ISSN 0016-7363

Measuring the Scale of Segregation Using *k*-Nearest Neighbor Aggregates

John Östh[1], William A. V. Clark[2], Bo Malmberg[3]

[1]Department of Social and Economic Geography, Uppsala University, Uppsala, Sweden, [2]University of California, Los Angeles, California, USA, [3]Department of Human Geography, Stockholm University, Stockholm, Sweden

Nearly all segregation measures use some form of administrative unit (usually tracts in the United States) as the base for the calculation of segregation indices, and most of the commonly used measures are aspatial. The spatial measures that have been proposed are often not easily computed, although there have been significant advances in the past decade. We provide a measure that is individually based (either persons or very small administrative units) and a technique for constructing neighborhoods that does not require administrative units. We show that the spatial distribution of different population groups within an urban area can be efficiently analyzed with segregation measures that use population count-based definitions of neighborhood scale. We provide a variant of a k*-nearest neighbor approach and a statistic spatial isolation and a methodology (EquiPop) to map, graph, and evaluate the likelihood of individuals meeting other similar race individuals or of meeting individuals of a different ethnicity. The usefulness of this approach is demonstrated in an application of the method to data for Los Angeles and three metropolitan areas in Sweden. This comparative approach is important as we wish to show how the technique can be used across different cultural contexts. The analysis shows how the scale (very small neighborhoods, larger communities, or cities) influences the segregation outcomes. Even if microscale segregation is strong, there may still be much more mixing at macroscales.*

Introduction

In recent years, studies of residential segregation have taken up the issue of scale in segregation studies. There have been important contributions by Wong (1993, 1999, 2004, 2005) and Reardon and colleagues who in a number of papers have explored the scale dimension of segregation (Reardon and Firebaugh 2002; Reardon and O'Sullivan 2004; Lee et al. 2008; Reardon et al. 2008, 2009). The scale of segregation is important when one considers different ways of measuring segregation, but can also be of importance when the effects and causes of segregation

Correspondence: John Östh, Department of Social and Economic Geography, Uppsala University, Box 513, SE-751 20 Uppsala, Sweden
e-mail: john.osth@kultgeog.uu.se

Submitted: March 26, 2013. Revised version accepted: April 24, 2014.

doi: 10.1111/gean.12053

are studied. Certainly, increased attention to modifiable areal unit problem (MAUP; Openshaw 1984) has been one stimulus for interest in the scale dimension of segregation. That the levels of segregation differ depending on what type of areal unit one uses to compute segregation indices is well known but is often not addressed.

In order to address the problem, Reardon et al. (2008) have proposed that segregation should be measured by constructing scalable egocentric neighborhoods for which the radius is allowed to vary in order to reflect different geographical scales. In this article, we provide support for the idea that scalable egocentric neighborhoods should be used in segregation measurement, but we propose that population size, instead of a radius, is used for measuring neighborhood scale. Technically, this is performed by expanding a buffer around each individual location until the population contained in the buffer reaches the threshold set for a specific scale level. This procedure will yield statistical aggregates for the *k*-nearest neighbors of different locations that can be used subsequently to compute different scale-dependent measures of segregation. Such *k*-nearest neighbor aggregates, called *bespoke* neighborhoods, were first introduced in studies of contextual effects (Johnston et al. 2000), but here we propose that they also provide an excellent tool for analyzing segregation patterns.

In this article, the usefulness of *k*-nearest neighbor aggregates will be demonstrated in a comparison of segregation patterns in three Swedish and one U.S. metropolitan region: Stockholm, Göteborg, Malmö, and Los Angeles. As we will note later, we believe that this technique can be used in different cultural contexts and largely solves the problem of different administrative units in different national contexts. One issue we will address is whether or not segregation levels are generally higher in large cities than in smaller cities. Krupka (2007), using local segregation data, argues that this idea is wrong. Our analysis shows that Krupka is essentially correct with respect to microscale[1] segregation but what distinguishes large cities from small cities is instead the proliferation of macroscale segregation which, we argue, will have distinct social and economic consequences from the microscale segregation that characterizes smaller urban areas.

The article represents, to our knowledge, the first attempt to compare segregation levels between urban areas in different countries using measures that are independent of administratively defined areal units. We would argue that such measures create an opportunity for better-designed comparative studies of both the causes and effects of residential segregation. The focus of the article, however, is not on the analysis of what mechanism lies behind the different type of segregation patterns, but rather to demonstrate the potential of this new approach.

For comparative research, there is not only the issue of the influence of different area subdivisions but also how we measure ethnicity and the ways of perceiving ethnicity and race. In this article, we use the standard racial categories for the Los Angeles case, but for Sweden we use a category based on national origin suggested by the Swedish National Board of Health and Welfare, namely, *visible minority*. (In Sweden, no racial categories are used by the national statistical office). In the European context, visible minorities consist of inhabitants' appearance, behavior, dress, habits, manners, religious customs, or way of speaking that is perceived as foreign by the majority population in each country (National Board of Health and Welfare 2010). This corresponds in Sweden to having your origin in a non-European, non-Anglo-Saxon country.

The visible minority is a concept developed in Canada but has been found to match the ethnic realities of contemporary Sweden at least as well as categories based on more narrow sets of national origin (see Tigervall and Hubinette 2010 and Molina 1997). The evidence for this comes from the close correlation between the location patterns of a narrowly defined group such as migrants born in sub-Saharan Africa and the location pattern of the entire visible minority group (see Appendix). There is also support for the notion of comparability from Pred (2000) who argues that Swedish views of immigrants are similar to racialized views of Blacks in the United Stated (p. ix). At the same time, we acknowledge that as comparative research develops there will be a need for more work on the comparative nature of ethnic and racial groups.

Segregation and scale

The established view in the literature is that the reason for changes in levels of segregation with the unit of analysis is that standard segregation measures are aspatial and each application reflects the particular geographic unit. That is, the values do not change if the spatial units are rearranged toward more or less clustering of homogenous neighborhoods. From a behavioral point of view, this suggests that no interaction takes place across the boundaries of the areal units and that we must use measures that are sensitive to a spatial rearrangement of units. A study by Jakubs (1981) amended the dissimilarity index by not only considering how many individuals would need to relocate to achieve ethnically balanced neighborhoods but also to what extent nonlocal relocation would be necessary. Morrill (1991) extended this discussion by proposing a different adjustment of the dissimilarity index, which takes into account the population composition of adjacent areas. These ideas are also embedded in the notions of distance and spatial association as outlined by Getis and Ord (1992).

A key reference with respect to the relationship between segregation and scale advances the field by relating the scale issue to the more general MAUP problem (Wong 1997). By comparing segregation measured using aggregates for block groups, census tracts, and towns, he demonstrates that segregation levels decline as one increase the scale of the areal units used for the measurement. To overcome the problems with aspatial segregation measures, Wong (1993) proposed that Morrill's method could be modified so as to reflect not only adjacency but also the length of shared boundaries and the compactness of neighboring areal units. He proposed the use of standard deviational ellipses for segregation measurement (Wong 1999), and in later measures ellipses that incorporate the population composition of increasingly distant adjacent areas (Wong 2005). In a later article with O'Sullivan and Wong (2007), he proposed that segregation measures can be based on kernel-based density estimates.

However, even if the scale dependency of segregation measures now is well understood, there has been less attention to the ways in which scale is of importance for the potential effects of segregation. Neither is there a full discussion of what mechanisms underlie macro- versus microscale segregation. This relative neglect of scale issues is exemplified by the population density invariance criteria that Reardon and O'Sullivan (2004) propose for segregation measures. This criteria implies that segregation values should be unchanged if the population density of different groups in a specific location is multiplied by a constant factor. Thus, a single neighborhood populated by 100 minority individuals and 10 majority individuals is equivalent to a neighborhood populated by 10,000 minority individuals and 1,000 majority individuals. This, effectively, implies that the scale of segregation is unimportant and, moreover, that the mechanism that fosters large- and small-scale segregation are essentially the same. In our view, this is a far too strong an assumption to make a priori. Instead, it is possible to advance different hypotheses about how the scale of segregation can affect outcomes (See Andersson and Malmberg 2014), and there is also a need for research on the scale dimensions of segregation processes.

The segregation profiles presented by Reardon et al. (2009) show segregation values based on aggregates of circular areas with different spatial extensions. Similarly, O'Sullivan and Wong

(2007) rely on spatial extension when they compute segregation values using bandwidths defined in kilometers. Wong (1997), on the other hand, defines scale on the basis of different types of geographical units that typically represent increasing population size, and uses aggregates of an increasing number of relatively equally populated census tracts to explore the scale of segregation in Washington DC (Wong 2005). Wong's use, in 1997 and 2005, of units linked to population size suggest that the scale of segregation can be viewed not only as spatial extension but also as linked to population size.

Important processes at the neighborhood level such as the formation of segregated schools, ethnically based organizations, and ethnically oriented businesses will be dependent on local size of the ethnic population. Moreover, it can be argued that varying levels of nearest neighbors (in our case the 100, 6,400, 51,000 nearest neighbors) fulfill different social roles as the people you are likely to encounter when you pick up your morning paper would consider inviting for a neighborhood barbecue, meet at the bus stop, meet as parents in the local school, or encounter at a local shopping mall. Thus, relating scale to population size can be as relevant as relating it to spatial extension.

Technically, computing aggregate statistics for neighborhoods defined by population size is more demanding than using spatially defined aggregates. But there are also advantages. With spatially defined areas, there is a risk that segregation values in cities with low population density will be inflated by random variance in population composition. If segregation values are based on equally populated areas, there will be no such effects of variations in population density. Using equally populated neighborhoods also eliminates the risk that an aggregate would in fact represent data for a single individual or household.[2] Segregation measures based on equally populated neighborhoods are therefore well worth exploring.

Methods and data

To show how individualized neighborhoods defined by population size can be used for segregation measurement, this article uses block-level population data from Los Angeles (U.S. Census 2010 data) and gridded register data from Sweden's three largest city regions (Statistics Sweden 2008). The data are used to create *bespoke* neighborhoods, which are neighborhoods of varying size using an individual's location and then computing the proportion of different groups of residents in the "bespoke" neighborhood. This is a different approach than that usually employed, whereby the first step consists of determining the spatial distribution of different population groups across a set of fixed areal subdivisions such as census tracts, moving windows techniques, or kernel-based density estimation.

In this article, the spatial distribution of different population groups is estimated by expanding a buffer around a given location until the population contained in the buffer reaches a threshold level. When the threshold is reached, the share of different population groups in the buffer population is determined and these figures are used to compute aggregate segregation statistics. Technically, the measure being proposed uses individual data that can be persons or households or some very small spatial unit such as a street block and then measures the probability of meeting another person (or aggregate of persons if it is a very small spatial unit) within some defined set of nearest people or nearest spatial units (Östh 2014; Östh, Malmberg, and Andersson 2014).

We use the isolation index to measure aggregate segregation across the bespoke neighborhoods. Our preference for the isolation index is based on behavioral considerations. From an individual point of view, segregation affects the probability of encountering people belonging to different subpopulations. What the isolation index provides is an aggregate measure of such probabilities. Given that we compute the population distribution for egocentric neighborhoods with a given population size, it is natural to choose the isolation index as the aggregate indicator of segregation. The population distribution of the egocentric neighborhood determines the probability of encountering representatives for different population groups if individuals are chosen randomly from the buffer population. The aggregate isolation index gives the same expected probability of meeting a member of a subgroup, but as a population group average.

That we prefer the isolation index as an aggregate measure does not mean that data on population composition based on equally populated, egocentric neighborhoods cannot equally well be used to compute entropy or dissimilarity indices of segregation. The entropy index, in many respects, is a useful tool for analyzing residential evenness of population groups. It does, however, lack an easily accessible intuitive meaning. The dissimilarity index certainly has an intuitive interpretation but, in comparison with the isolation index, is less easy to link to theories about the effects of segregation. Therefore, in this article where the focus is on analyzing scale effects on segregation, we see no strong arguments against using the isolation index as our aggregate measure.

Using individualized neighborhoods to compute interaction probabilities has several advantages. First, contrary to measures that depend on statistical areas that may differ in size, using individualized neighborhoods provides an opportunity to construct measures that are computed in exactly the same way across different urban areas. Second, by using different population threshold, segregation can be analyzed at different scales. Third, it provides an opportunity for statistical offices to provide detailed geographical information on contextual variables without endangering confidentiality. Finally, the use of individualized neighborhoods offers a measure that has theoretical appeal by focusing directly on the link between individuals and their behaviors and their environments.

To compute the probability of an individual of one group meeting individuals from the same or a different group, we use a software (Östh 2014) that makes it possible to find the *k*-nearest neighbor (using a variety of *k*-values between 50 and 409,600 persons) of each individual or small populated unit in data sets containing millions of locations and very large numbers of people. The software calculates the share of individuals belonging to a user-specified subgroup for each *k* (user-defined count of surrounding individuals) at each populated location in the studied area. The share can then be used to describe exposure (interaction or isolation) at *k* various scales on the block level.

When the data set is very large, such as occurs for city blocks in large metropolitan regions, computing *k*-nearest neighbors is a large-scale undertaking. To make this analysis feasible, we transform the geographical information from block centroid coordinates to a disaggregated grid. In the greater LA data set used in this article, all blocks in the entire five county region were geographically matched to a grid having a spatial resolution of 250 × 250 ft. This means that all geographical "finer-than-grid" specifications will be rounded to the nearest 250 × 250 ft point in the grid. Compared with using original block centroid coordinates, gridded data have the advantage of enabling automated and fast yet precise computation of nearest neighbor—this is because in gridded space any next nearest location is always located at the same distance and relative location from any location of origin (as squares on a chessboard). This is also the reason why very populous and large regions may be included in the analysis. The preparation (data gridding) of U.S. Census data for analysis in the EquiPop process that produces the index is illustrated in

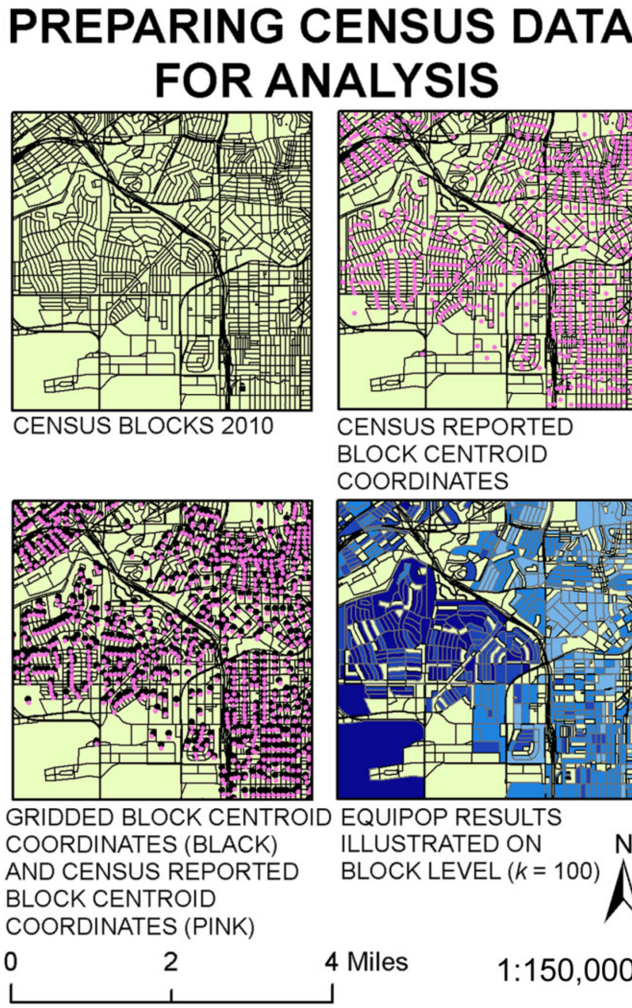

Figure 1. Gridding of U.S. Census data. Upper left illustrate the geographical distribution of blocks in a randomly selected area of Los Angeles. The upper right section illustrates the distribution of populated census that reported block centroid coordinates. The lower left illustrates the overlap of census that reported populated block centroid coordinates (pink) and the corresponding gridded (250 × 250 ft) points colored black. The lower right section illustrates that EquiPop output can be exported and illustrated on a block polygon level as long as the point grid is finer than the polygon data.

Fig. 1. Because the finest geographical level of Swedish statistics offered for research already is rectified to a grid of 100 × 100 m, no further preparation of the Swedish data is necessary.[3]

For each populated point in the gridded data, two population values are needed; first, the local total sum of individuals and second, the local total sum of individuals belonging to a subgroup. The subgroup population must always be part of the total population. The EquiPop software uses cumulative counts of local total sums of people collected from the surrounding next nearest neighbors. Whenever the cumulative count of surrounding people reaches a user-defined k-nearest neighbor value, the share of individuals belonging to the subgroup is saved. The overarching rule of population counting in the EquiPop software is to add the next nearest population. This means that counting starts locally and continues to count the populations from adjacent, not yet counted locations. Consequently, the output produced by the program always represents the closest neighbors from each location of origin. The *k*-values are used to define the population within which contacts of various kinds can occur. That is within 100 nearest neighbors how many subgroup individuals are potentially capable of meeting, or in other words how many of the same group are likely to be exposed to others of a similar background. In the substantive presentation that follows, we examine *k*-values limits of 100 (a local neighborhood), 6,400 (a community), 51,000 (a small city), and 400,000 (a metropolitan region).[4]

Spatial isolation (SI) with a population threshold *k* is computed as

$$Spatial\ Isolation_k = \frac{\sum_{i=1}\left(x_i * \frac{x_{i,k}}{k}\right)}{\sum_{i=1}(x_i)}$$

where *k* represents a predetermined count of nearest neighbors and x_i the size of the minority population in block *i*. The formulation $\frac{x_{i,k}}{k}$ represents the share of the minority population among the *k*-nearest neighbors of individuals living in block *i*.

The analysis proceeds in two steps. First, the output from the EquiPop program is used to compute the block-level probability of encountering individuals from specific ethnic groups among the nearest 100, 6,400, 51,200, and 409,600 neighbors. Second, the probability values are used to calculate SI index values for the different ethnic groups and metropolitan areas for the different *k*-levels. SI index values are aggregated and illustrated on tract (United States) and SAMS (Sweden) level in the Results section.

Most studies of segregation have been nationally specific, but there is increasing interest in how segregation varies by cultural context. The tools developed here allow a nearly comparable approach with measuring segregation in different contexts. However, as in most instances, the data available will differ in important ways. In the context for this study, U.S. and Swedish statistics differ in a number of ways—censuses have been discontinued in Sweden and replaced with longitudinal, annually updated registers containing public records for all residents. Specifically, the Swedish data do not report measures of population ethnicity. How then can we compare issues of segregation in such contexts with those in the United States where the focus on segregation has been the front and center?

To create comparability, we can split the Swedish population into two groups based on their country of birth. Individuals born in Africa, Asia (excl. Russia), and the Americas south of United States are listed as belonging to a "visible minority." All others are categorized as "majority population." The classification is simple in the sense that Sweden-born children to visible minority migrants are listed as majority population members—and similarly, all individuals born in Africa, Asia, and Latin America are listed as visible minorities regardless of ethnic background. However, because immigration from non-European countries is a relatively recent phenomenon and because most recent immigrants face increased risk of unemployment, welfare dependency, and poor health, these groups are at greater risk of being victims of discrimination. Thus, they are comparable with groups (Blacks and Hispanics in particular) who have been subject to discrimination in the United States. Comparing isolation of different ethnical groups in LA to a broadly defined group of immigrants in Sweden makes sense in that both categorizations define groups as being minorities and majorities in their regional context.

To address the question of strict comparability of ethnic and racial groups, we provide in the Appendix the data for sub-Saharan Africans in the cities in Sweden. These data are nearly directly comparable with the data for Blacks in Los Angeles, which gives us the opportunity to compare two nearly comparable groups—a subsection of the visible minorities (Africans) in Sweden with Blacks in Los Angeles. The results suggest that even though there are differences, they are small and more comparable than different. We also note our earlier reference to the research by Pred (2000) on ethnic comparability.

Naturally, we are measuring populations in their residence, and the issue of nighttime and daytime populations is of growing interest in measuring segregation. That analysis is beyond this article, but we note that using workplace data it would be possible to measure what are essentially daytime levels of segregation. There are a number of problems to be overcome before such an analysis can be completed, but the use of individual measures as proposed here is a solid way to conduct this analysis.

Results

The maps of the central area of the Los Angeles metropolitan area demonstrate the varying levels of SI across both ethnicity and space (Fig. 2). It is useful to focus initially on the $k = 6{,}400$ maps as that k-level represents what we can think of as the extended neighborhood, about the size of the census tract in the U.S. context. For African-Americans, there are two areas where the probabilities of meeting other African-Americans are relatively high (over 0.8), but these small concentrations seem minor compared with the extensive area in which the likelihood of meeting another African-American is in the range of 0.3–0.5 and even more extensive areas where the probabilities are less than 0.25. At the other extreme of the ethnic patterns, those of the Asian population show that the probabilities of meeting other Asians never rise much above 0.5 and for significant parts of the city the probabilities of an Asian meeting another Asian are less than 0.1. Both Asians and African-Americans then, overall, are relatively well integrated into the metropolitan landscape in the overall patterns. This is shown quite clearly at the larger scales that we have called the city scale at $k = 51{,}000$.

The Latino population is now a majority of the city of Los Angeles and close to a majority in the county. Thus, for much of the central area of Los Angeles Hispanics have high probabilities of meeting other Hispanics. Only in the peripheral areas are the probabilities of meeting other Hispanics low. It is in the areas of low probabilities of any of the visible minorities meeting other visible minorities where we find the areas of greatest concentrations of the fourth large population group—Whites. The outcome reflects the base distribution of the White population, which has the highest probabilities of meeting other Whites in the coast areas in these maps.

The results for the Swedish metropolitan areas are shown in Fig. 3. To facilitate comparison of isolation between Los Angeles and the Swedish cities, SI value intervals and corresponding colors in both Fig. 2 (LA) and Fig. 3 (Sweden) are identical. We should also note that the cities in Sweden are structurally more similar to Los Angeles[5] than might appear at first sight. Los Angeles is quite similar in density to Stockholm (about 7,000 persons per square mile compared with about 9,000 per square mile in Stockholm).

The three Swedish cities (Fig. 3) display patterns of SI of visible minorities that are similar for especially the lower k-values. Isolation of visible minorities is localized to densely populated, multistory building suburbs in the outskirts of the urban areas. Surrounding areas have relatively few visible minority individuals. As the k-values increase, the clustering of visible minorities

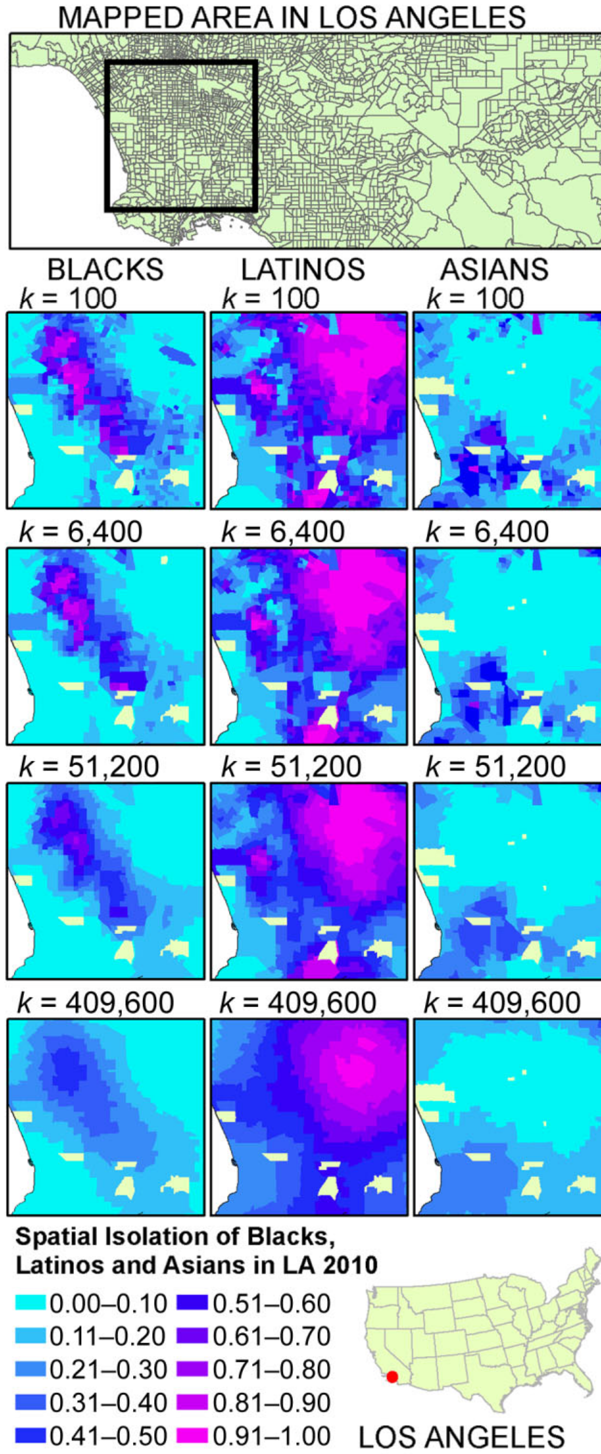

Figure 2. Probability of encountering Black, Hispanic, and Asian individuals among the nearest 100, 6,400, 51,200, and 409,600 neighbors in Central Los Angeles.

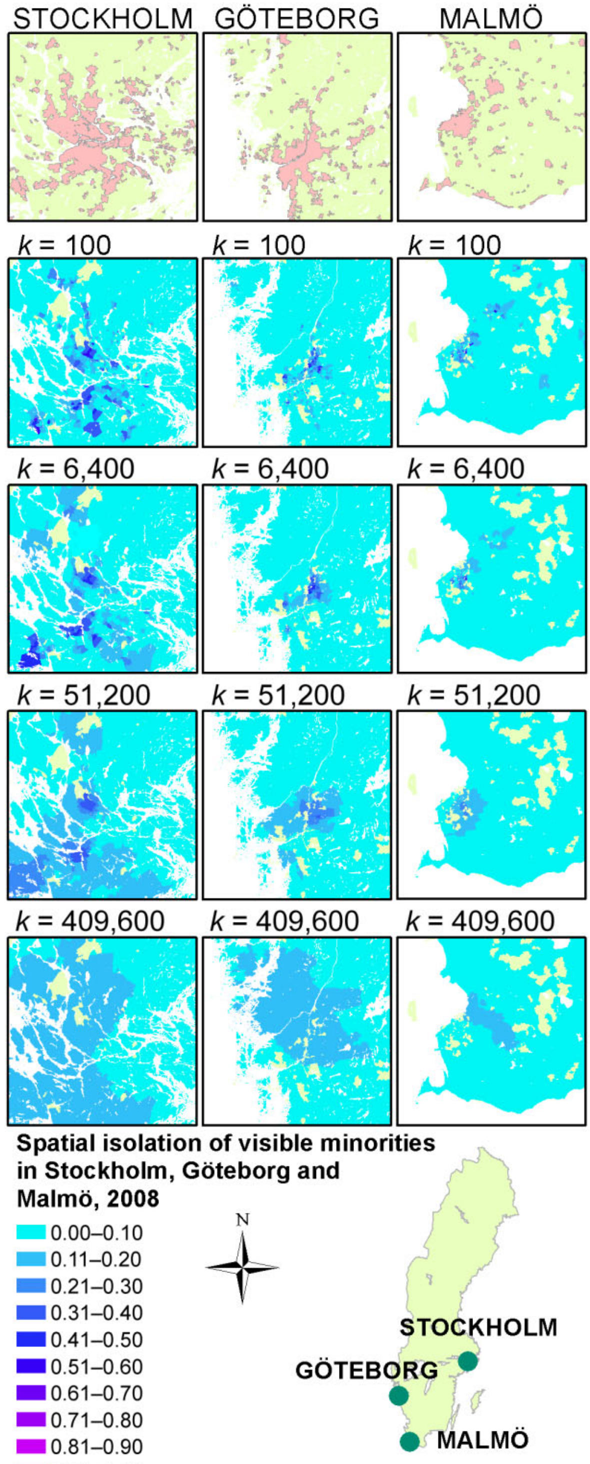

Figure 3. Probability of encountering visible minority individuals among the nearest 100, 6,400, 51,200, and 409,600 neighbors in the Stockholm, Göteborg, and Malmö metropolitan region. Top row shows distribution of urban and rural areas.

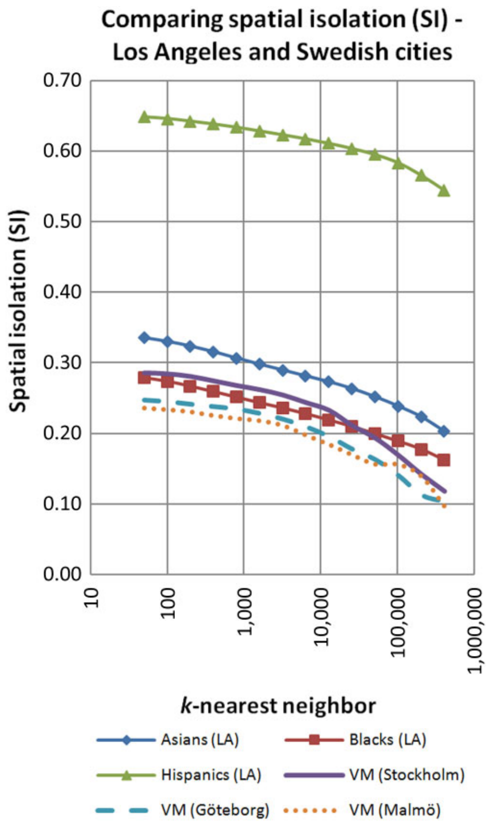

Figure 4. Spatial isolation in Los Angeles and in Swedish cities. Values on the *y*-axis represent spatial isolation for different groups (LA) or cities (Sweden). Values on *x*-axis represent log-scale values of *k*-nearest neighbors. VM (Swedish cities) represents visible minority.

"spills" over to adjacent areas, reducing the differences between high-isolation areas and few visible minority individual areas. Especially in the Western parts of Stockholm, SI values over 0.5 in a few locations and over 0.4 in others indicate that isolation of foreign-born visible minorities exists also in city large structures (k = 51,200). At k = 409,600 Stockholm is divided into a Western more visible minority clustered region and an eastern coastal and affluent region. Similar patterns, yet on a smaller scale, are detectable around the other Swedish cities as well as in Los Angeles—coastal areas are predominately confined to non-minority groups.

How do these maps improve our understanding of segregation and isolation? Low *k*-levels are useful to denote segregation on a "dog-barking" distance, grouping individuals using the same bus stop, recognizing each other as neighbors, or shopping at the same store. Larger *k*-levels are useful to denote regions wherein large shares of one's lifetime activities are contained. Large scales may therefore be useful to identify regions or contexts in which the population is born, goes to school, works, marries, votes, identifies with a sports team, joins religious communities, and does so as being in the center of a minority region, or far from any minority individuals, or anywhere in between.

Fig. 4 shows aggregated SI index in scaled graphs that capture the changing levels of segregation as the count of nearest neighbors changes. Values in Fig. 4 clearly show that the isolation of Hispanics exceeds isolation among all other studied groups regardless of *k*-value. This is a direct effect of two conditions: a large number of Hispanics in LA and a clustering of Hispanics in certain areas in LA (see Fig. 2). As described in Fig. 4, the isolation of the average Hispanic person in Los Angeles County is profound at any chosen scale—at local scales, the average isolation means that two-thirds of all individuals close to a Hispanic individual are of Hispanic origin too. Even at very large scales, the average isolation is always greater than 0.5.

For Asians and Blacks in Los Angeles, and for visible minorities in Sweden, isolation proves to be more similar. Although SI of Asians is greater at all *k*-levels compared with Blacks and visible minorities in Sweden, the macroscale *k*-level isolation of Asians (0.2 at $k = 409{,}600$) is lower than the isolation of all groups at microscale *k*-levels (>0.2 for all at $k = 50$). This indicates that although isolation varies between *k*-levels and groups, studied groups (almost always) experience the same levels of isolation at different scales.

Apart from the deviating patterns of Hispanic isolation, there are notable similarities and dissimilarities in the Los Angeles material and the Swedish material that deserve attention. The fact that isolation of visible minorities is almost the same in Sweden as for the comparator groups in LA may be surprising to many, but the strong concentration of certain ethnic groups in some Swedish municipalities has already led to international media attention (see for instance, Washington Post 2008). In fact, the level of SI among visible minorities in Stockholm is very slightly greater than for Blacks in Los Angeles for all *k*-levels below 51,200 (in Appendix we note that for sub-Saharan Africans the indices are lower than for visible minorities). The difference between Los Angeles and the Swedish cities lies instead on how isolation develops at regional scales (larger *k*-levels). The SI of the visible minority population in Sweden drops more rapidly at *k*-levels of 51,200 and greater compared with the corresponding trend in Los Angeles.

Conclusions and discussion

An important conclusion in the literature on spatial measures of segregation is that segregation is a scale-dependent phenomenon yet most indices do not specifically address the scale issue. In this article, we have proposed an SI measure that uses individualized scalable neighborhoods with fixed population size to measure spatial variations in population composition. With this approach, the scale of segregation is linked to the number of people that live in the neighborhood for which the population composition is estimated. We argue that for segregation measurement, population size is a relevant measure of scale for the same reason that size of a city, in general, is measured by its population count and not by its spatial extension. Moreover, in the same way as city-size distributions are used to compare the urban structure of different countries, segregation measures based on individualized neighborhoods with the same population count can be used to compare segregation levels for across urban areas and across countries.

The usefulness of the proposed approach has been demonstrated using data for the Los Angeles and for three Swedish metropolitan areas. Like earlier studies, our results show that segregation measured as the probability of encountering a same-group individual in the neighborhood declines with increasing neighborhood scale (or alternatively of encountering a different individual, increases). Our data have also provided some support for the idea that segregation in large metropolitan areas is more macroscale than in smaller metropolitan areas. This can, as we see it, have implications for how segregation will influence the urban process. If there is only microscale segregation, social interactions that are linked to high-level services are more likely to involve different ethnic groups than in areas where there is macroscale segregation. In areas with macroscale segregation, it is, instead, possible for high-level services to be directed toward specific ethnic groups.

It is, of course, no surprise that macroscale segregation is absent in small metropolitan areas. However, it would be perfectly possible that also large metropolitan areas were characterized by microscale segregation. But microscale segregation is not what we find in the Los Angeles case when it is compared with the Swedish metropolitan areas.

Research on spatial segregation during the last 20 years have brought into existence an extensive toolbox of segregation measures, but there are still no clear signs of what will become the standard approach adopted by a majority of the researchers. Instead, outside a circle of specialists, there seems to be a continued reliance on aspatial measures of segregation. This is an unfortunate situation in view of the widely acknowledged deficiencies of the aspatial measures. One possibility of progress in this area could, however, be that more attention is given to the way different measures of segregation can help us explain how segregation affects individuals and communities, or the extent to which different measures of segregations could help to distinguish between processes that lead to higher or lower levels of segregation. This would imply that different measures would not be evaluated in isolation. Instead, a certain measure becomes interesting to the extent that it can be used to formulate successful hypotheses of causal links or empirical regularities.

Notes

1 Following a recommendation by two reviewers, we use micro and macro rather than small and large scale to identify segregation that occurs locally versus that which occurs at community- and city-wide scales.
2 We are aware that the number of mixed-race individuals and mixed-race households has increased in the United States and a second article considers the mixed-race population as a special case.
3 100 m roughly equals 328 ft.
4 A website with the EquiPop software is available from url: http://equipop.kultgeog.uu.se.
5 Despite popular perceptions of Los Angeles as a sprawling city, the population densities in the city are the highest of metro areas in the United States.

Appendix. The spatial distribution of sub-Saharan Africans and visible minorities in Swedish cities

In this article, we claim that it is reasonable and theoretically of interest to compare the spatial isolation of the visible minority group in metropolitan Sweden with the spatial isolation of acknowledged racial groups in the Los Angeles area. The justification for this claim is that processes of ethic segregation are not, in general, based on objective measure such as skin color but on how ethic and racial categories are perceived in a specific national context. Below we present data that support our claim that the visible minority group in the Swedish context is a racial/ethnic category comparable with Black, Hispanics, or Asians in the LA context.

We support our claim by showing that a subcategory of the visible minority group, sub-Saharan Africans, in spite of making up only about 10% of the visible minority population in Sweden has a spatial distribution that is very similar to the spatial distribution of the entire visible minority group but very distinct from the spatial distribution of the majority population (Table A1, A2, and A3). If segregation processes in the Swedish context would have singled out the Black group as distinct from Asians or Hispanics, such great similarities in the spatial distribution would have been highly improbable.

Table A1 Correlation Coefficients (Pearson) between Local Counts of Africans and VM, Africans and Poor, and Africans and Individuals Born in Sweden

	Africans and VM	Africans and born in Sweden
100 × 100 m	0.803**	0.359**

** Correlations are significant on 99% level. All of Sweden.
VM, visible minorities.

Table A2 Correlation Coefficients (Pearson) between Differently Populated *k*-Neighborhoods of Africans and VM, Africans and Poor, and Africans and Individuals Born in Sweden

	Africans and VM	Africans and born in Sweden
k 100	0.617**	–0.428**
k 200	0.662**	–0.480**
k 400	0.697**	–0.525**
k 800	0.724**	–0.559**
k 1,600	0.742**	–0.582**
k 3,200	0.749**	–0.594**
k 6,400	0.766**	–0.610**

** Correlations are significant on 99% level. All of Sweden.
VM, visible minorities.

Table A3 Correlation Coefficients (Pearson) between Differently Populated *k*-Neighborhoods of Africans and VM, and Africans and Individuals Born in Sweden for Stockholm, Göteborg, and Malmö

	Stockholm	Göteborg	Malmö
k 100	0.790**	0.719**	0.590**
k 200	0.816**	0.777**	0.640**
k 400	0.842**	0.807**	0.715**
k 800	0.867**	0.839**	0.774**
k 1,600	0.882**	0.869**	0.834**
k 3,200	0.894**	0.896**	0.862**
k 6,400	0.912**	0.929**	0.899**

** Correlations are significant on 99% level.
VM, visible minorities.

In addition, Table A4 shows that sub-Saharan Africans are not a strongly segregated group as evidenced by the isolation index, whereas the segregation of the visible minority group is substantial. These results corroborate the claim made by, for example, the Swedish National Board of Health and Welfare that in Sweden it is the visible minority group as such that is influenced by segregation pressures.

Table A4 Spatial Isolation (SI) for Africans, Visible Minorities, and Individuals Born in Sweden—for Locations: Sweden, Stockholm, Göteborg, and Malmö, *k*-Values 100 and 6,400

		Sweden	Stockholm	Göteborg	Malmö
Africans	*k* 100	0.089	0.130	0.102	0.047
	k 6,400	0.036	0.099	0.066	0.026
Visible minorities	*k* 100	0.207	0.287	0.255	0.242
	k 6,400	0.145	0.250	0.219	0.206

References

Andersson, E. K., and B. Malmberg. (2014). Contextual effects on educational attainment in individualised, scalable neighbourhoods: Differences across gender and social class. Urban Studies:0042098014542487.

Getis, A., and J. K. Ord. (1992). "The Analysis of Spatial Association by Use of Distance Statistics." *Geographical Analysis* 24, 198–206.

Jakubs, J. F. (1981). "A Distance-Based Segregation Index." *Journal of Socioeconomic Planning Sciences* 15, 129–41.

Johnston, B., K. Jones, S. Burgess, C. Propper, R. Sarker, and A. Bolster. (2004). "Scale, factor analyses, and neighborhood effects." *Geographical Analysis* 36(4), 350–68.

Krupka, D. J. (2007). "Are Big Cities More Segregated? Neighbourhood Scale and the Measurement of Segregation." *Urban Studies* 44(1), 187–97.

Lee, B. A., S. F. Reardon, G. Firebaugh, C. R. Farrell, S. A. Matthews, and D. O'Sullivan. (2008). "Beyond the Census Tract: Patterns and Determinants of Racial Segregation at Multiple Geographic Scales." *American Sociological Review* 73(5), 766–91.

Molina, I. (1997). "Stadens rasifiering: etnisk boendesegregation i folkhemmet." Doctoral, Uppsala universitet, Uppsala, Sweden.

Morrill, R. L. (1991). "On the Measure of Geographical Segregation." *Geography Research Forum* 11, 25–36.

National Board of Health and Welfare. (2010). "Social rapport 2010." Stockholm, ISBN 978-91-86585-00-6.

O'Sullivan, D., and D. W. S. Wong. (2007). "A Surface-Based Approach to Measuring Spatial Segregation." *Geographical Analysis* 39(2), 147–68.

Openshaw, S. (1984). "The Modifiable Areal Unit Problem." Concepts and Techniques in Modern Geography #38 Norwich England, Geo Books.

Östh, J. (2014). Introducing the EquiPop software – an application for the calculation of k-nearest neighbour contexts/neighbourhoods. Available from url: http://equipop.kultgeog.uu.se

Östh, J., B. Malmberg, and E. Andersson. (2014). "Analysing Segregation with Individualized Neighbourhoods Defined by Population Size." In *Social-Spatial Segregation: Concepts, Processes and Outcomes*, 135–161, edited by C. D. Lloyd, I. Shuttleworth and D. Wong. Bristol: Policy Press.

Pred, A. (2000). *Even in Sweden: Racisms, Racialized Spaces and the Popular Geographical Imagination*. Berkeley, CA: UC Press.

Reardon, S. F., and G. Firebaugh. (2002). "Measures of Multigroup Segregation." *Sociological Methodology* 32, 33–67.

Reardon, S. F., and D. O'Sullivan. (2004). "Measures of Spatial Segregation." *Sociological Methodology* 34, 121–62.

Reardon, S. F., S. A. Matthews, D. O'Sullivan, B. A. Lee, G. Firebaugh, C. R. Farrell, and K. Bischoff. (2008). "The Geographic Scale of Metropolitan Racial Segregation." *Demography* 45(3), 489–514.

Reardon, S. F., C. R. Farrell, S. A. Matthews, D. O'Sullivan, K. Bischoff, and G. Firebaugh. (2009). "Race and Space in the 1990s: Changes in the Geographic Scale of Racial Residential Segregation, 1990–2000." *Social Science Research* 38(1), 57–72.

Tigervall, C., and T. Hubinette. (2010). “Adoption with Complications: Conversations with Adoptees and Adoptive Parents on Everyday Racism and Ethnic Identity.” *International Social Work* 53(4), 489–509.

Washington Post. (2008). Iraqi Refugees Find Sweden’s Doors Closing, written by Mary Jordan, April 10, http://www.washingtonpost.com/wp-

Wong, D. (1999). “Geostatistics as Measures of Spatial Segregation.” *Urban Geography* 20(7), 635–47.

Wong, D. (2004). “Comparing Traditional and Spatial Segregation Measures: A Spatial Scale Perspective.” *Urban Geography* 25, 66–82.

Wong, D. W. (2005). “Formulating a General Spatial Segregation Measure.” *Professional Geographer* 57(2), 285–94.

Wong, D. W. S. (1993). “Spatial Indices of Segregation.” *Urban Studies* 30, 559–572.

Wong, D. W. S. (1997). “Spatial Dependency of Segregation Indices.” *The Canadian Geographer* 41, 128–36.