KULTURGEOGRAFISKT SEMINARIUM

Multiscalar Typology of Residential Areas in Sweden

Juta Kawalerowicz

Bo Malmberg

2021:1

Stockholms universitet

Multiscalar Typology of Residential Areas in Sweden

Juta Kawalerowicz

Bo Malmberg

ISBN PDF
ISSN 0347-9552

Distributör: Kulturgeografiska institutionen, Stockholms universitet, Stockholm

Content

Figures and Tables

Preface

This report is based on a project "The Neighbourhood Revisited: Spatial polarization and social cohesion in contemporary Sweden" directed by Bo Malmberg at the Department of Human Geography at Stockholm University. The project is funded by the Swedish Foundation for Humanities and Social Sciences (Riksbankens Jubileumsfond, RJ), grant registration number M18-0214:1 Between 2020-2021 Statistics Sweden carried out a survey of residents living in different type of neighborhood context and the multiscalar typology presented in this report was used to develop a sampling for this survey. A dataset containing coordinates for inhabited grid cells together with their cluster assignment has been deposited on Stockholm University research repository - figshare.com and can be downloaded from there. Earlier versions of this work have been presented to our colleagues at Human Geography Department and Stockholm University Demography Unit, we are thankful for their comments and feedback.

Stockholm, May 2021
Juta Kawalerowicz and Bo Malmberg

Sammanfattning

I denna rapport presenterar vi en geografisk indelning av grannskap som bygger på en analys av hur befolkningens sammansättning varierar när hänsyn tas både till vilka som bor i det omedelbara grannskapet (de 200 närmaste grannarna), de som bor i omgivningar som sträcker sig över större områden (de närmaste 51 200 grannarna), liksom till grannskap i storlekar däremellan (som omfattar de närmaste 400, 800, 1 600, 3 200, 6 400, 12 800 och 25 600 grannarna).

Denna typologi har utvecklats för att besvara forskningsfrågor som som formulerats i projektet ”Lyckliga gatan?”, nämligen i vilken utsträckning individs attityder tar form i och påverkas av rumsliga kontexter. De variabler vi använder avser situationen 2016 och bygger på individuella registerdata som tillgängliggjorts av Statistiska Centralbyrån och de mäter såväl socioekonomisk som etnisk segregation. Dessa data innehåller statistiska uppgifter om befolkningen i 250 meters-rutor (i tätorter) och 1000 meters-rutor (utanför tärorter), och befolkningssammansättningen i omgivningen beräknas utifrån mittpunkten av rutor som är befolkade. I våra data finns det sammanlagt 213 663 befolkade rutor. För att identifiera platser som har samma typ av geografisk kontext har vi använt oss av en hierarkisk clusteranalys som succesivt parar ihop de platser som är mest lika. Vi har till slut valt en indelning med tio olika typer av områden.

Vår analys ger en bild av vilka olika typer av grannskap som finns i Sverige 2016 och hur de skiljer sig åt när det gäller inkomstnivåer, sysselsättning och etnicitet, när hänsyn tas till sammansättningen på olika skalnivåer. Ett viktigt resultat är att det inte bara är i större städer som man hittar områden men en hög koncentration av utlandsfödda utan sådan områden finns också utanför storstadsområdena på mindre orter. En skillnad jämfört med storstäderna är dock att det här är frågan om småskalig segregation men på samma sätt som i storstäderna är detta områden med många fattiga. Den etniska mångfald som har funnits i storstäderna återfinns nu alltså även utanför dessa. Ett annat viktigt resultat är att storstäderna kännetecknas av stora kontraster inte minst i inkomstnivåer, och att det finns stora geografiska skillnader i befolkningens utbildningsnivå.

Summary

In this report we present a typology of residential areas based on multi-scalar measures of population composition computed for individualized neighborhoods with equal population size. This typology has been developed to address one of the research questions proposed in "The Neighbourhood Revisited" project namely to what extent individuals' attitudes are formed in and influenced by their residential context. Our indicators come from 2016 register data and measure the extent of socio-economic and ethnic segregation at nine scales ranging from 200 to 51,200 closest neighbors. We use hierarchical clustering methods to develop a typology where we are able to assign each of 213,663 residential areas (i.e. inhabited grid cells) in Sweden into one of ten cluster types.

Our analysis offers a data-driven insight into what are the most typical residential contexts in Sweden in 2016 and how they are distinguished in terms of socio-economic affluence, labor market attachment and immigration at different scales. One novel finding is the emergence of rural town diversity, where areas characterized by presence of non-Western immigrants at small scales are no longer confined to the largest cities. Our analysis also shows that education and high income are the key indicators that drive differentiation between urban and rural cluster types.

1. Introduction

The aim of this analysis is to explore the spatial sorting of individuals based on their socio-demographic characteristics. The result will show how strong the tendencies are for individuals with similar characteristics to congregate in space and also what type of locations different groups are found in. Theoretically, this type of analysis can be linked to a tradition initiated by the Chicago school of sociology which, in the early 20th century explored the spatial differentiation of Chicago as this city experienced a rapidly growing population based on migration as well as rapid economic development and growing social inequality. To some extent, the situation in Sweden in the early 21st century has similarities to Chicago in the early 20th century. Sweden as well has experienced large inflows of migrants, a growing economy, and increasing inequality. This makes it interesting to explore the kind of spatial differentiation that has resulted.

In analyzing the socio-demographic structure of Chicago, the leading members of the Chicago school (Park 1926, p. 3) looked for inspiration in vegetation geography and, more specifically, a book published in 1895 by a Danish botanist, Eugen Warming: "*Plantesamfund* - Grundtræk af den økologiske Plante-geograf", later translated to German (1896), and English (1909) with the title "*Oecology of Plants—an introduction to the study of plant-communities"*. Based on this book, Warming has been heralded as the founder of Ecology. What Warming does in this book is to systemize the findings made during the 19th century concerning the spatial distributions of different plants across the world and the factors that influence which plants are found together in different locales. The Chicago school has sometimes been criticized for importing concepts from the natural science into sociology but, as can be seen from the title of Warming's book, the importation of concepts also goes the other way. Warming's analysis of plant geography was inspired by sociological concepts such as *samfund* and communities.

One important aspect of Warming's book is that it was based on the descriptive results of extensive field works by different botanist. This report fulfils a similar role as those descriptive studies. It will describe spatial patterns as a basis for considerations of what factors influence spatial sorting.

The interest in the factors that drive spatial sorting is not the only rational for this report. Spatial sorting results in neighborhoods with different socio-demographic composition. This is of interest because neighborhoods composition will influence the way neighborhoods work: the consumption patterns of your neighbors, if your neighbors are resourceful or hard-pressed economically, their cultural interest, their norms and ambition regarding education, and possibly also their world outlook, political preferences, and other values. In the literature, there is evidence that neighborhood composition can influence the life choices that individuals make (for a recent overview, see Graham 2018). These studies often focus on how life course outcomes in early adulthood are influenced by neighborhood context during adolescence (Chetty and Hendren 2018, Chetty et al. 2014), a period during which individuals life expectations are formed and during which individuals have substantial contacts with peers that are from the same neighborhood.

In order to determine what indicators to use in the analysis we have, on the one hand, started out from earlier studies of neighborhood factors that can influence individual level outcomes (Malmberg and Andersson 2019, Andersson and Malmberg 2018, Andersson and Malmberg 2015, Wimark, Haandrikman and Nielsen 2019) and, on the other hand, studies of segregation patterns (Malmberg et al. 2016, Malmberg and Clark 2020, Nielsen and Hennerdal 2019, Costa et al. 2018).

Studies of how neighborhood composition influences individuals emphasize high income levels, the level of education, indicators of low economic status and, to a lesser extent, indicators of ethnicity. In the segregation literature there is a division between studies of ethnic segregation and studies of socio-economic segregation.

Strong patterns of ethnic or racial segregation have been observed in the United States, especially between blacks and whites but, are common also in most urban areas that have experienced large migrant inflows (Andersson, Lyngstad and Sleutjes 2018). According to a well-established theory, migrants at arrival will chose to reside in areas where there is a high concentration of their own ethnic group (Vogiazides 2018). One reason is that this helps to overcome difficulties associated with language skills, and that it will be easier to develop social networks in neighborhoods where members of your own nationality live. But ethnic segregation can also be driven by an interest in avoiding contact with other ethnic groups (Clark 1991).

Socio-economic segregation is often seen as driven by high-income groups who, because of their purchasing power, will be able to outbid other groups in the housing market (Bischoff and Reardon 2014). High-income group also has a strong tendency to congregate in geographical space. One reason for this could be that urban areas have location that in terms of their landscape, access to natural amenities and environmental quality are highly valued. But it could also be that living in high-income areas will signal status, or that living there will help you to develop valuable social contacts, and will help your children develop skills, contacts, and attitudes that will give them a head start in their educational and occupational careers (Toft 2018). Similar considerations can contribute to the spatial concentration of groups with a higher educational attainment.

Housing market factors can also contribute to a spatial concentration of individuals with low income and low purchasing power (Andersson, Wimark and Malmberg 2020). Low purchasing power will tend to sort poor individuals into areas that have a low valuation because of poor housing standards, poor environmental quality, less attractive landscape features, or locations that are located far away from jobs and have poor communications. In this way, spatial sorting may perpetrate concentrated the concentration of poverty and lead to ever higher rates of unemployment. Another factor is a desire to avoid low-status ethnic groups is association with potential loss in valuation of housing in areas with high concentration of minorities.

Studies of how neighborhood composition influences individuals emphasize, to a large extent, similar indicators: high income levels, the level of education, indicators of low economic status and, to a lesser extent, indicators of ethnicity. Previous research guides our choice of indicators. Our indicators can be divided into indicators of socio-economic affluence, socio-economic deprivation, and migration history (see the selection for in ResSegr project, for details see Haandrikman et al. 2019). One argument for looking at both socio-economic affluence and deprivation is that including just one indicator could obscure our classification. For instance, some affluent neighborhoods with higher shares of older residents may look like low-income areas if our classification was based solely on income. Additionally, in border areas some residents commute to work and are not registered with Swedish authorities. Yet, having no income in Sweden may not mean that they are poor, in this case it is necessarily to consider other socio-economic indicators, for instance reliance on social assistance. Next, we look at different domains of socio-economic affluence and deprivation. Even though income affluence tends to go hand in hand with social status, there are differences between professional elites and business elites, which our analysis is able to pick up. By coupling socio-economic indicators with immigration, we can show how socio-economic segregation and ethnic segregation are intertwined. Additionally, by adding immigration, our analysis reveals that neighborhoods with high concentration of immigrants are no longer characteristic of urban areas.

The Chicago School account of natural areas

The Chicago School account of the spatial differentiation of urban areas, best elucidated by Harvey W. Zorbaugh (1927) in his contribution to the Ernest W. Burgess volume “The Urban Community”, starts out from a generalization about typical patters in the expansion of American cities, leading to the establishment of a Central Business District, a zone of transition, a zone of working men’s home containing sectors of rooming-house districts, a zone of apartments and single family dwellings and, finally, a commuter zone of sub-urban areas. Cutting through these zonal arrangements are belts of industries along railroads, lines of local transportations which can generate retail business streets, leading also to sectoral differentiation of the city. Zorbaugh then concludes that “this framework of transportation, business organization and industry, park and boulevard systems, and topographical features … break the city up into numerous smaller areas, which we may call natural areas, in that they are the unplanned, natural product of the city's growth” (p. 222). Where “railroad and industrial belts, park and boulevard systems, rivers and rises of land acting as barriers to movements of population tend to fix the boundaries of these natural areas” (ibid), with their centers gravitating towards intersections of business streets. Each of these areas “acquires a physical individuality” based on their location relative to business, industry, and transportation as well as in relation to natural features, the influence of which will show up in land values or rent levels.

According to Zorbaugh, this almost physical structuring of the city is the basis for a spatial sorting of the population. Zorbaugh’s idea about this sorting is that individuals in a city stand in a competitive relation to each other, a competition for economic positions which also includes a competition over physical locations: “In this competition for position the population is segregated over the natural areas of the city”. Exactly how this happens is, however, not detailed:” Land values, … tend to sift and sort the population”. ”Cultural factors also play a part … creating repulsions and attractions”. The result, on the other hand, is clear:

And as a result of this segregation, the natural areas of the city tend to become distinct cultural areas as well a "black belt" or a Harlem, a Little Italy, a Chinatown, a "stem" of the "hobo," a rooming-house world, a "Towertown," or a "Greenwich Village," a "Gold Coast," and the like—each with its characteristic complex of institutions, customs, beliefs, standards of life, traditions, attitudes, sentiments, and interests. The physical individuality of the natural areas of the city is re-emphasized by the cultural individuality of the populations segregated over them. Natural areas and natural cultural groups tend to coincide. (p. 223).

Zorbaugh, here, also points out a correspondence to plant ecology:

That is, just as there is a plant ecology whereby, in the struggle for existence, like geographical regions become associated with like "communities" of plants, mutually adapted, and adapted to the area, so there is a human ecology whereby, in the competition of the city and according to definable processes, the population of the city is segregated over natural areas into natural groups. And these natural areas and natural groups are the "atoms" of city growth, the units we try to control in administering and planning for the city. (pp. 223-224)

Today, neighborhood or community have been substituted for the term natural area which in understandable given that the connotations of "natural" are unclear. What also could have been lost, however, is the idea about how neighborhoods are formed through a process of spatial sorting. Here it could be argued that the plant ecology metaphor still could be of value, with physical conditions including transportation infrastructure as a basis, upon which a sorting of individuals is imposed, and with the resulting population composition influencing future decision about staying, moving out, or moving to the neighborhood. Moreover, composition of the neighborhood also will have consequences for what type of community that develops. It is also possible to add, though, that an ecological theory of the neighborhood still is present even if no direct references are given to Chicago school ideas. One example is Schelling's theory of ethnic preferences as a determinant of processes of segregation (1971). Building on time-geography, Allan Pred (1984) envisions the production of places in terms that can be seen as inspired by the Chicago school, and the same could be said of Kirsten Simonsen (2016). A final example is Robert Sampson, whose theory of the neighborhood more clearly is built on the foundation of Chicago school ideas (2012).

Social area analysis and factor ecology

During the interwar years, urban analysis in the Chicago school tradition was engaged in neighborhood studies that, to a large extent, were based on qualitative studies. It was only after 1940 that is was possible to see a strong development in census based studies of urban differentiation. Here the pioneering study was Eshref Shevky and Marilyn Williams (1949). They studied the Los Angeles area and constructed three indices intended to capture important structuring dimensions. The first index measures socio-economic status—social rank—and was based on occupational data, education data, and rent levels as measure of income. The second index measured family status using fertility levels, female employment, and proportions of single-family houses as indicator variables. The third index measured ethnic segregation using the location quotient for five different ethnicities: Blacks, Mexicans, Asians, Russians, and Italians. Mapping of these indexes demonstrated that areas with low social rank were found adjacent to Los Angeles' manufacturing zone, and areas with high social rank in environmentally pleasant areas in west Los Angeles and in Pasadena. The family status index showed a clear center periphery pattern with low fertility, high female employment, and few single-family dwellings in the center. Areas with high levels of segregation overlapped with low-status areas. In their report, the selection of these indices is motivated by considerations of social trends, but it has been questioned if the indexes instead were inductively derived. It could also be argued that they represent abstractions of Chicago school observations. Clearly, Park, Burgess and others did report spatial sorting based on social status and ethnicity, but they also observed differences in family status related to a concentration of boarding housed in central locations. Note also that the spatial assimilation to some extent can be seen as a family cycle, with young migrants initially settling in central areas and later moving out from the central areas to single-family housing areas in less central locations. Here, it can be noted that in the 1940s, the US baby boom had not yet started and, thus, many young women had not yet formed families. If these women lived in central Los Angles this could be reflected in high female employment rates and low fertility. In a follow-up study, Bell (1953) was able to replicate Shevky and Willimas's study using data for the San Francisco Bay area, and two years later Shevky and Bell (1955) published a much cited summary and a guide on how to conduct social area analysis. Originally, social area analysis was based on indexes designed to capture dimensions that were hypothesized to be important. Here, an important step forward was taken by Bell (1955) when he demonstrated that it was possible to use factor analysis for assessing if, in fact, the proposed dimensions were sufficient for capturing the urban spatial differentiation. The introduction of multivariate statistical methods for the analysis of urban data led to the establishment of a very active research field that came to be known as factor ecology. Research in this field was stimulated by readily available data from censuses that provided a wealth of information on the level of census tracts. The late 1950s and the 1960s was also a period when computers increasingly were provided university researchers with the ability to process numerical data in more efficient ways. Moreover, factor ecology could be seen as providing a possibility for finding generalizable empirical patterns in the spatial structuring of urban areas in a way that would realize some of the expectations of scientific progress that can be seen as immanent in the human ecology approach of the Chicago school.

Factor ecology can be seen as a relatively successful approach (for an overview see Janson 1980). In general, broadly in line with the Shevky and Williams or the Shevky and Bell results, it was possible to show that indicators of census tract composition could be summarized by a relatively small number of common factors, and these factors often were related to the dimensions proposed in the original studies.

A weakness of the factorial approach, however, is that it is based on the correlation of different aggregate indicators measured at the level of census tracts. One reason for such correlations can be correlations at the individual level. For example, if families with children have a tendency to live in single-family dwellings, a sorting of such families into certain areas will result in a correlation between the average number of children and the proportion of single-family homes. Another example could be a correlation between high income and having a long education. But correlations between census-tract measures could also result if different types of individuals are sorted into the same areas. For example, it could be the case that both young people in their early 20s and people above age 70 are found in areas with many small apartments. This implies that the correlation matrix of ecological variables may represent a mixture of processes, and this mixture will be reflected in the type of factors that result from the analysis. That is, even if common factors can be used to describe the spatial differentiation found in urban areas, it is not guaranteed that the factors can be linked in a direct way to the underlying processes of spatial sorting. There is a risk that the factors become somewhat diffuse, maybe suggestive not always with a clear interpretation.

In the field of factor ecology, different ways of solving these problems were tried, but it cannot be safely concluded that these solutions were successful. It was not enough just to add more data, and neither could the problems be overcome by exploring different statistical approaches.

The inability to completely solve these problems is likely to have contributed to a decline in the popularity of factor ecology from the 1980s and onwards (Butler and Hamnett 2012). This decline is understandable but not an unmixed blessing. It is still true that patterns of spatial differentiation can be captured by restricted number of common factors, and it is still true that these factors are suggestive of underlying processes of spatial sorting. This can be seen as a reason for a continued use of factor ecology as an exploratory method. Factor ecology can be used to provide a generalized picture of spatial sorting, and thus point to what type of processes that need to be studied in more detail. Even if one should not forget that factor ecology in itself is not sufficient for a full understanding (Janson 1971).

Swedish studies

Factorial ecology was introduced in Sweden in the 1960s (Westergaard 1970). There are however relatively few international publications that report results from such studies (Janson 1971, Aldskogius 1982). One highly cited publication is (Wikstrom 1991). Here a factorial model is estimated for Stockholm in order to extract neighborhood characteristics that can be of importance for crime patterns. The resulting factors correspond in broad terms to the Shevky-Williams-Bell model with a familism factor, as social rank factor, and a factor that captures migrant density overlapping with precarity. Dahlbäck (2016) can be seen as a follow-up study to Wikstrom’s analysis. Apart from crime studies, the factorial approach has also been taken up in health studies as a tool for capturing neighborhood characteristics that are of importance for health outcome. These studies, however, tend to concentrate on a single dimension that captures social deprivation (Sundquist, Malmstrom and Johansson 2004). One study that in some respect is more similar to original social area analysis is Biterman and Franzén (2007). They classify neighborhoods in Stockholm, Malmö, and Gothenburg both along an ethnic dimension and with respect to a socio-economic dimension, and then analyze how these dimensions are related.

In the 2010s the factorial approach has been revived in Swedish research that measures geographical context using individualized scalable neighborhood instead of census tracts (Swedish SAMS or DeSO neighborhood units). With this approach it is acknowledged that spatial context is not something that characterizes a geographical subarea. Rather, context should be seen as relative to a specific location. Moreover, with an increasing access to individual, geocoded register data, and the increasing processing power of computers it has become feasible to identify the neighboring population of specific locations, and to assess composition of this neighboring population in terms of relevant indicators. Such individualized neighborhoods, also called ego-hoods or bespoke neighborhoods, have the advantage of being scalable. One can choose to assess the composition of the 100 nearest neighbors, the 800 nearest neighbors, the 6,400 nearest neighbors, or other neighborhood scales that are deemed to be relevant. That is, it becomes possible to have multi-scalar measures of geographical context. In relation to the exploration of spatial sorting this implies that sorting processes that operate on different geographical scales can be analyzed. An early application of this approach is (Malmberg, Andersson and Bergsten 2014). Here, eight different indicators are used (unemployment, tertiary education, single motherhood, high income, newly arrived migrant, social allowance, no employment, and non-European migrant) using scale levels ranging from the nearest 12 neighbors to the nearest 12,800. A factor analysis of this data yielded 15 common factors according to the conventional eigenvalue larger than one criterium. The large number of factors needed to describe the data can be related to the use of multi-scalar measures. Thus, some of the factors obtained by (Malmberg et al. 2014) have high loadings only for small-scale measures of context, some represent contexts that extend across spatial scales, and a third group capture context that represent the population composition of areas that are removed from the closest neighborhood. One important finding in this study is that there is an elite factor working across scales that dominates the structuring of the social landscape. These large-scale elite concentrations are primarily found in metropolitan areas. In contrast, small scale elite areas tend to be found in smaller urban areas. A later study (Andersson, Abramsson and Malmberg 2019) shows that the spatial structure captured by these multi-scalar common factors can be linked to different housing preferences.

Other studies in the same tradition are Andersson and Malmberg (2015), Andersson and Malmberg (2018) and Malmberg and Andersson (2019). All these find a large-scale elite factor that dominates the spatial structure. By including a housing variable, they find a factor that can be interpreted as the familial factor. In keeping with the Shevky-Williams-Bell tradition they also find migrant-density factors that overlaps with precarity, as in Wikstrom (1991).

A somewhat different perspective is taken in (Andersson et al. 2020, Malmberg, Andersson and Wimark 2018). Here, it is tenure type mixing in neighborhoods that is in focus. That is, these studies single out tenure types as having an influence on spatial sorting. In the 2018 study, the authors show that the multiscalar approach makes it possible to see the influence of tenure type mixing policies on the spatial structure of urban areas in Sweden, especially outside the metropolitan areas. What is found here is a concentration of a single tenure type at small scale levels but with increasing mixing at larger scales. But there are also more homogenous tenure type landscapes where a single tenure type is dominant across different scale levels. The 2020 study instead shows how specific tenure type landscapes are linked to specific patterns of social and ethnic segregation.

Taken together, these studies, on the one hand, shows that the factorial ecology when applied to multi-scalar contextual data is helpful for capturing important dimensions in the spatial structure of geographical contexts. What is added by the multi-scalar approach is the understanding that spatial sorting does not occur on specific geographical scale but on varying geographical scales. If a weakness of traditional factorial ecology is that variation across census tracts captures on-going processes of spatial sorting only in a diffuse way, this is addressed with the multi-scalar approach by giving hints about the spatial scale of the sorting processes. Moreover, it has been demonstrated that scale matters when it comes to the effects of variation in geographical context.

In addition to factor ecological studies there have also been many studies of segregation patterns, often with a focus on metropolitan areas or only on Stockholm. For an overview of such studies see Vetenskapsrådet (2018). While these studies documented ethnic and income segregation and the association between deprivation and segregation in urban context, it is possible that they may have overlooked emerging patterns of segregation in smaller Swedish towns.

2. Methods and data

In this chapter we discuss the methods used for our classification. Readers more interested in the results can go directly to the next chapter and come back to this chapter in order to understand how the analysis has been carried out.

To classify residential areas based on their population composition it is common to apply multivariate cluster analysis techniques. Researchers often use fixed geographical sub-divisions such as Census tracts, yet results of such analyses are strongly influences by how boundaries of such sub-divisions are drawn (Openshaw 1984). Additionally, by focusing exclusively on such geographical sub-divisions, it is assumed that wider context (outside of such units) does not matter. Using aggregates for individualized scalable neighborhoods to measure context has been proposed as a method for addressing this problem (Östh et al. 2014). Such individualized neighborhoods can be constructed by expanding a buffer around different residential locations until the population encircled by the buffer corresponds to a selected population threshold. When this threshold is reached, one can compute aggregate statistics on selected socio-economic variables for the encircled population. By varying the population threshold, contextual measures computed in this way can be designed to focus only on the closest neighbors or on a larger number of neighbors. In the present study, we allow k to vary from 200 to 51,200 in successive doublings of the population thresholds. The computation was carried out using a Geocontext software, developed by Pontus Hennerdal (2019).

Geocontext requires that the input data be geo-coded to a high level of detail. For this we used Migrant Trajectories database at Stockholm University. This database contains register-based, individual level information for the population in Sweden from 1990 to 2016 with geocodes of the residential location in 250-meter squares for urban and 1,000-meters for rural areas. Seven individual level indicators were extracted to use as input for Geocontext: (1) Having a tertiary education, (2) Having taxable income in the highest decile, (3) Being in employment, (4) Having received social allowance during the year, (5) Being at risk of poverty, (6) Country of birth outside of Sweden in EU/EFTA country, (7) Country of birth outside of Sweden in non-EU/EFTA country. Table 1 provides a description of variables and how they were coded. The individual level data was aggregated to grid cell squares or residential areas based on their geo-coordinates. In a small number of cases (mostly in sparsely populated locations) 250-meter squares were not available, then data was aggregated at 1,000-meter squares.

Table 1 Variables

Variable	Description
Tertiary education	Share aged 25-64 with tertiary education
High income	Share aged 25-64 who have levels of taxable income in the highest decile
Employment	Share aged 25-64 in employment
Social assistance	Share aged 18-64 who have received social assistance at some point in the reference year
At risk of poverty	Share of persons who are 25 or older who have disposable income below 60 percent of the median disposable income value
non-European immigrants	Share born outside of EU28 regions or Nordics and no parents born in Sweden
European immigrants	Share born in one of the EU28 regions or Nordics and no parents born in Sweden

Using 7 indicators at 9 scales (k = 200, 400, 800, 1600, 3200, 6400, 12800, 25600 and 51200) gave a total of 63 measures of neighborhood context that can be used to classify residential areas using cluster analysis. However, since many of these variables will be highly correlated, we used factor analysis that compresses the 63 original indicators to 8 orthogonal factors before proceeding with the cluster analysis (for a similar method see Clark et al. 2015). The reason for initial factor analysis is that given the strong correlation between indicators across different scales it is possible to capture most of the variation in neighborhood composition using a small number of factors. Moreover, given the number of our observation, reducing the number of measures of neighborhood context from 63 to 8 makes clustering algorithms more computationally manageable. The factor analysis was based on correlations and the number of factors was selected based on them having eigenvalues higher than one. The factors were rotated using the varimax method.

Figure 1 to Figure 8 show the results of the factor analysis. The panels in these figures show the loading of the different factors for each indicator. Figure 9 to Figure 16 show geographic distributions of quintiles for each factor, where 1 means *“very low”* and 5 *“very high”* value for a given factor. Note that it is possible for one residential area to have “very high” value on more than one factor. The first factor represents the *large-scale elite context* with high factor loadings for tertiary education and high income. Figure 9 shows that this factor is concentrated in large cities. The second factor represents *small scale disadvantage and diversity* because of high values for social assistance and non-EU/EFTA immigrants at lower k-levels (closer neighbors). Figure 10 shows that this factor is finer grained than *large-scale elite context* and more evenly spread out. The third factor represents *large-scale diversity* (high factor loadings at all k-levels). Figure 11 shows that it operates at larger scale than *small scale disadvantage and diversity* and tends to have low values in mid-Sweden. The fourth factor is a mirror reflection of factor two but at larger k-scales, suggesting that is represents areas *adjacent to disadvantage and diversity*. Figure 12 shows that it is common for suburban areas. Factors five and six signal *adjacent to small-scale disadvantage* and *adjacent to small-scale diversity* with high values for disadvantage and diversity for medium range k-values. The main difference is that for factor five we also observe high factor loadings for non-EU/EFTA immigration, while for factor six it is less pronounced. Figure 13 and Figure 14 reveal that the former picks up more fine gained differences (this is also visible in Figure 5 and Figure 6 where we see that the peak in values for disadvantage and diversity is more to the left for *Adjacent to small scale disadvantage*). The seventh factor represents *large-scale poverty;* we also note that for this factor disadvantage appears to be unrelated to immigration. The eighth factor represent *large-scale high income* with high levels of top income earners at all k-scales that, unlike for *large-scale elite context*, are not coupled with high educational attainment. Figure 15 and Figure 16 show that in areas where values tend to be high for *large-scale poverty* they tend to be low values for *large-scale elite context* and vice versa, although there are some areas which have high values for both factors (for example in Vellinge municipality in Skåne which is located in the south-western tip of Sweden)

Figure 1 Large scale elite

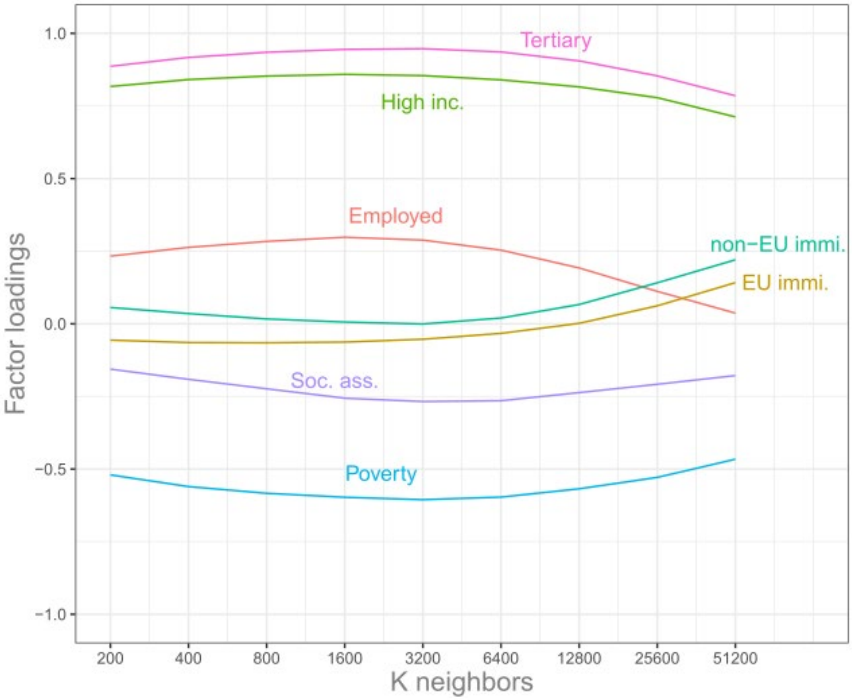


Figure 2 Small scale disadvantage & diversity

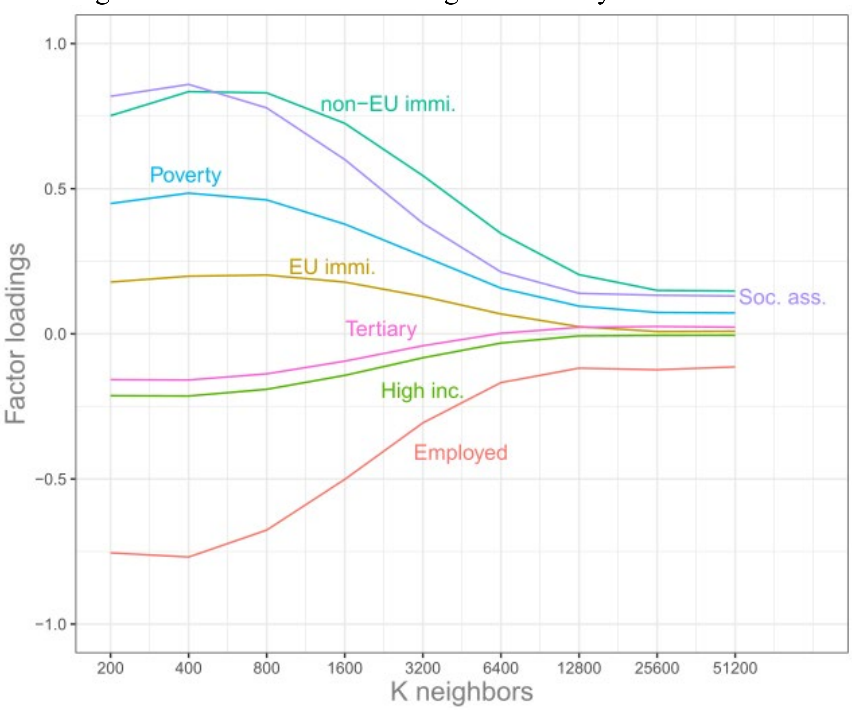


Figure 3 Large scale diversity

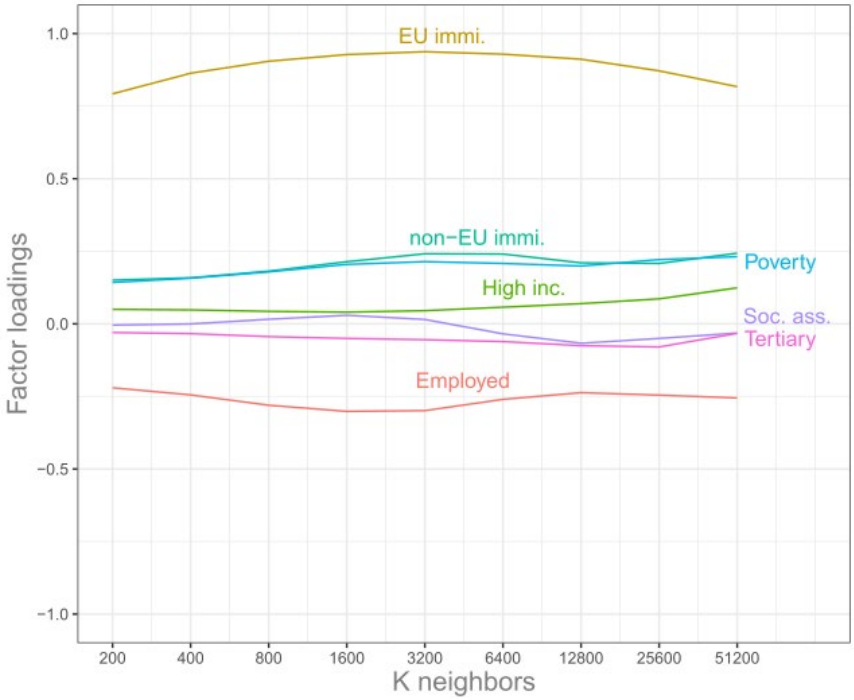


Figure 4 Adjacent to disadvantage & diversity

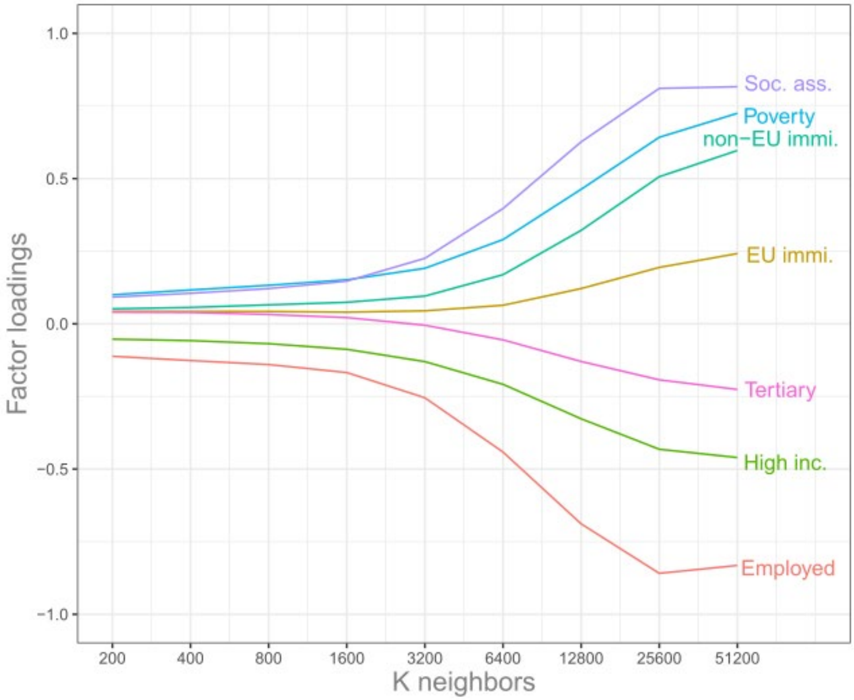


Figure 5 Adjacent to small scale diversity

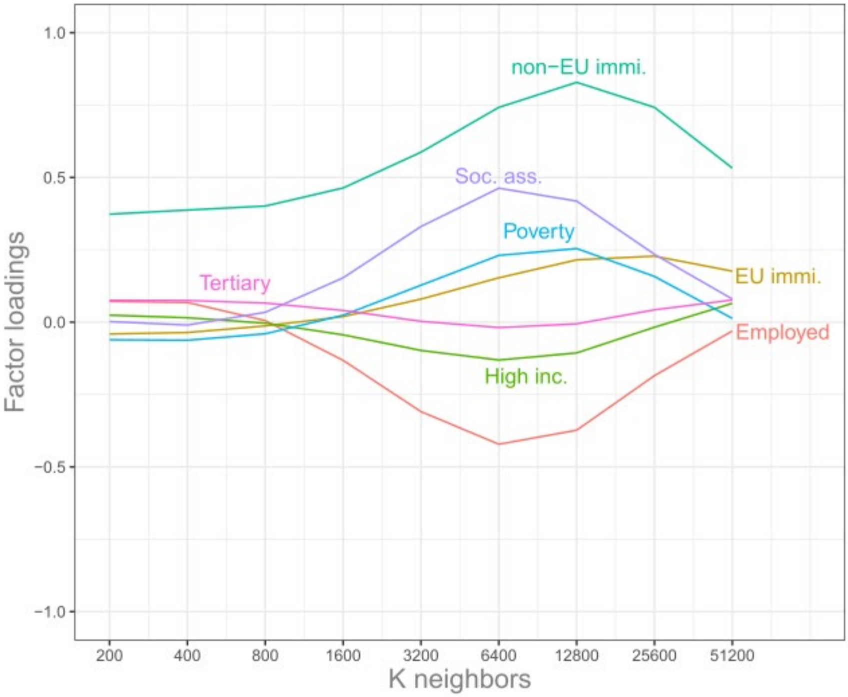


Figure 6 Adjacent to small scale disadvantage

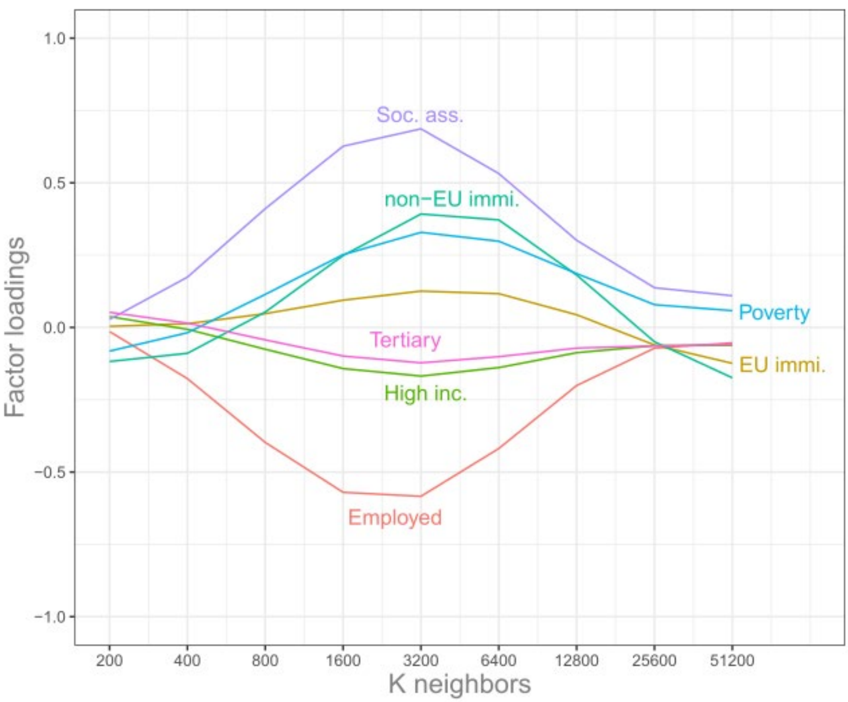


Figure 7 Large scale poverty

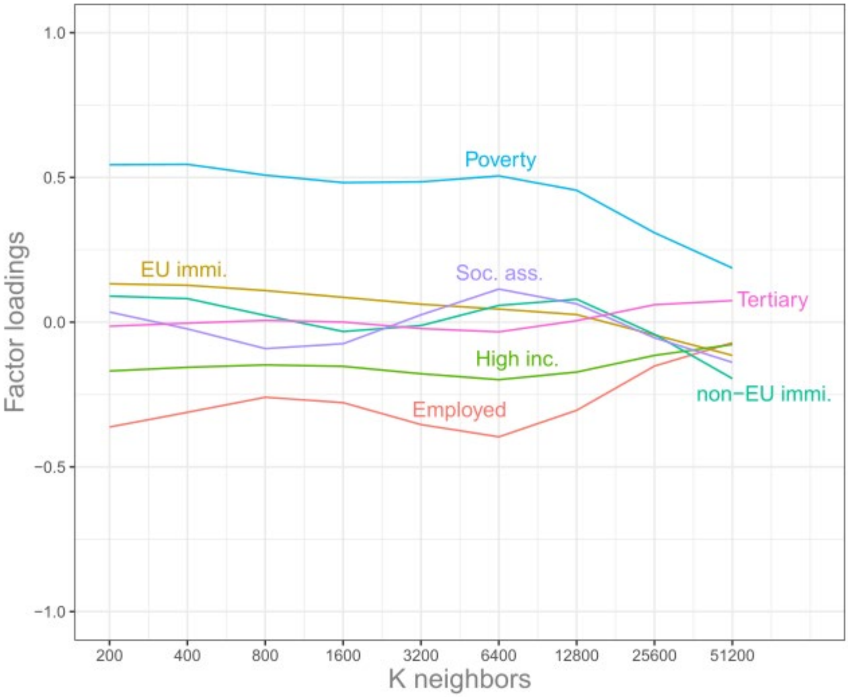


Figure 8 Large scale high income

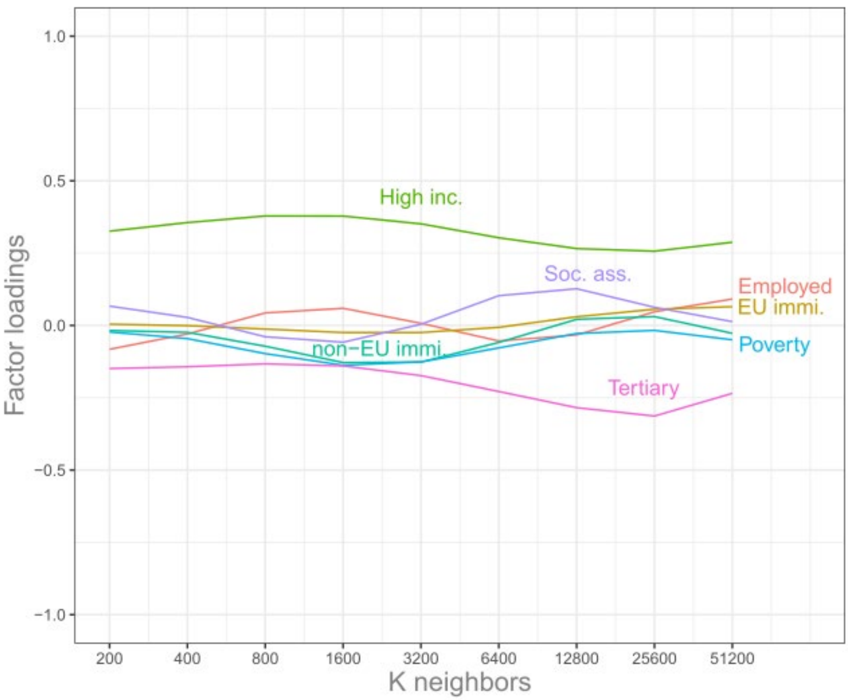


Figure 9 Large scale elite, geographical distribution

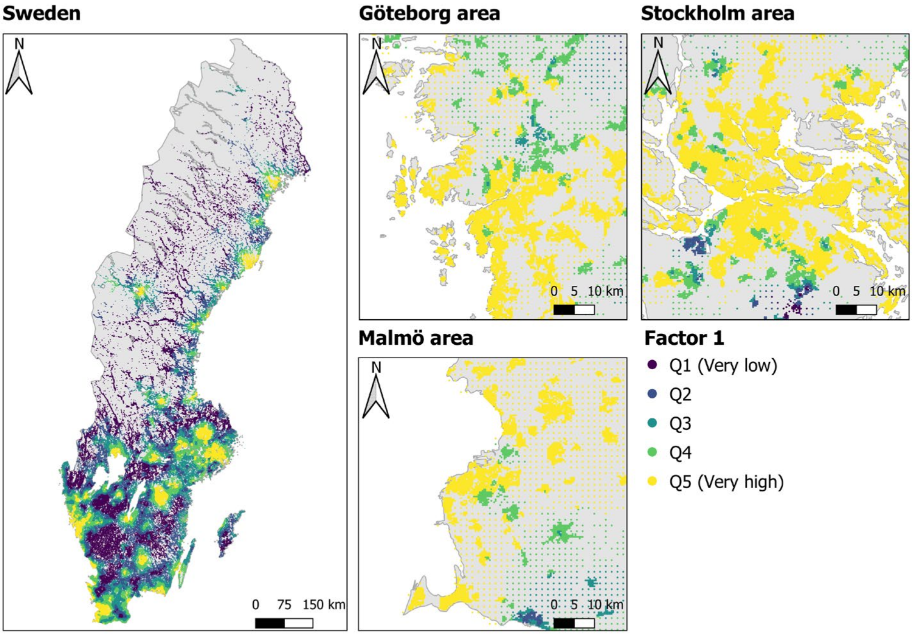


Figure 10 Small scale disadvantage & diversity, geographical distribution

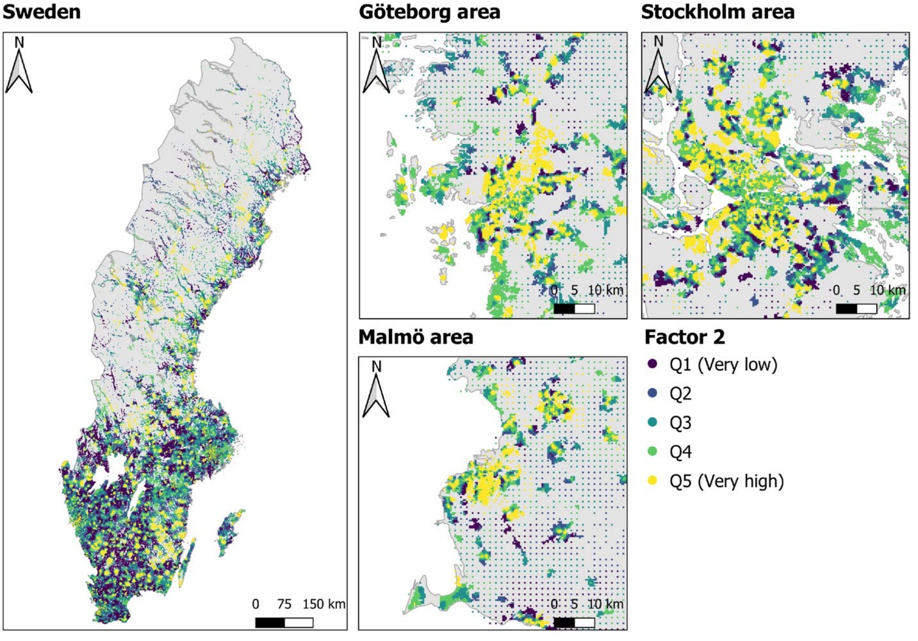


Figure 11 Large scale diversity, geographical distribution

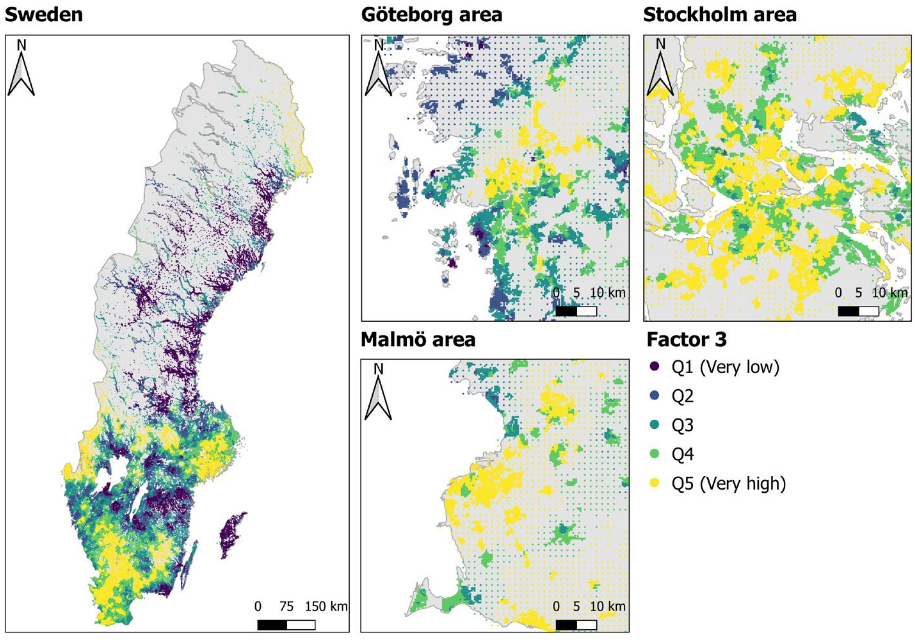


Figure 12 Adjacent to disadvantage & diversity, geographical distribution

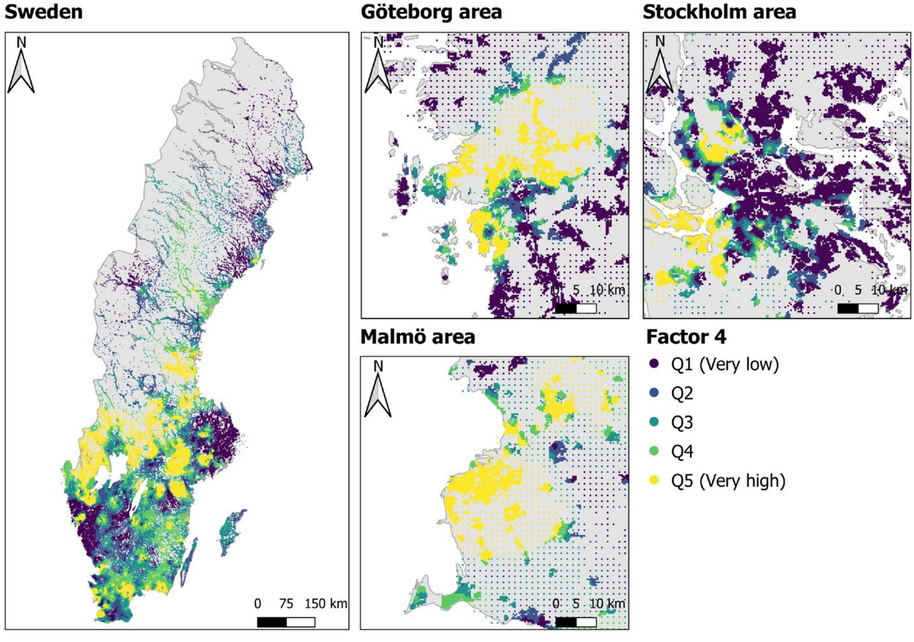


Figure 13 Adjacent to small scale diversity, geographical distribution

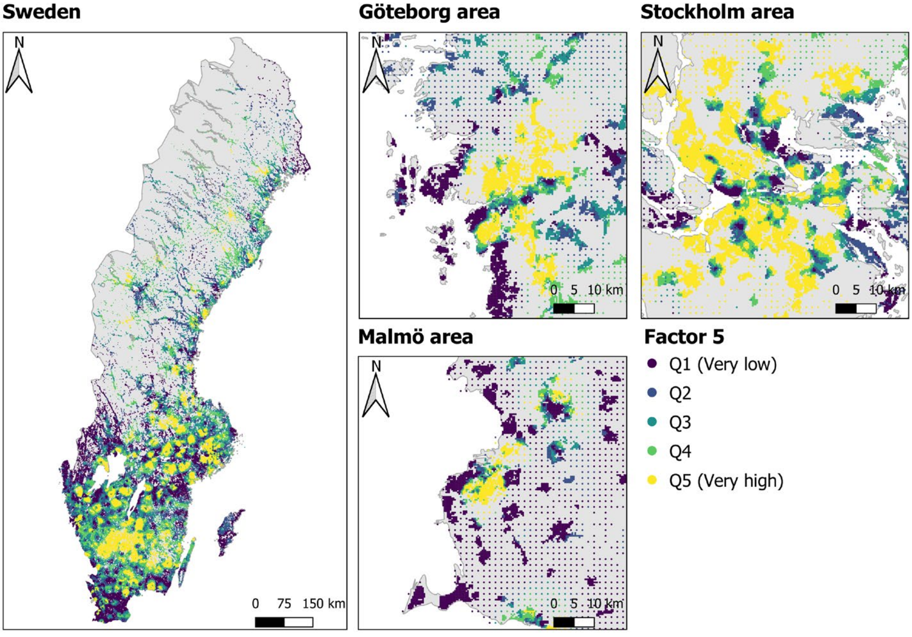


Figure 14 Adjacent to small scale disadvantage, geographical distribution

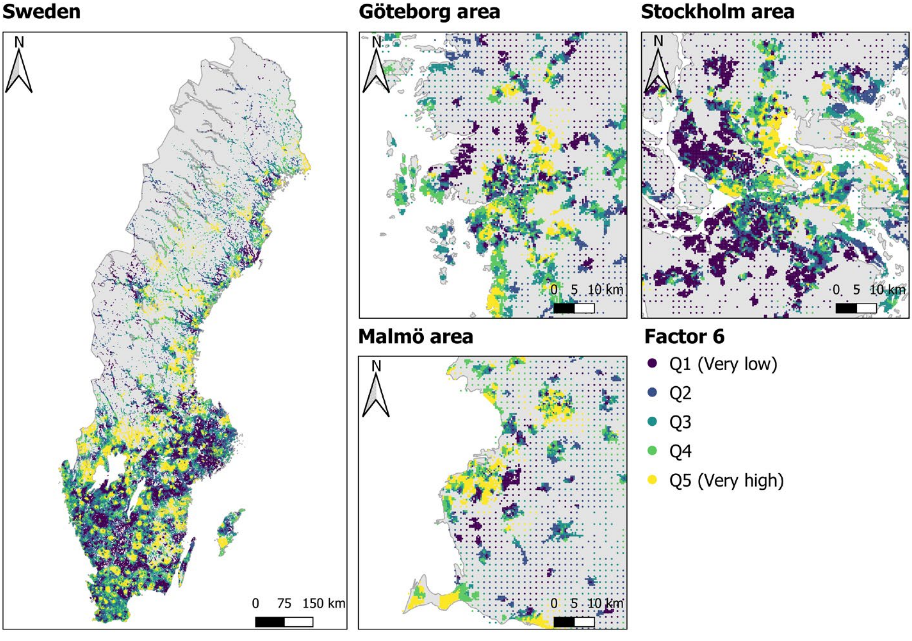


Figure 15 Large scale poverty, geographical distribution

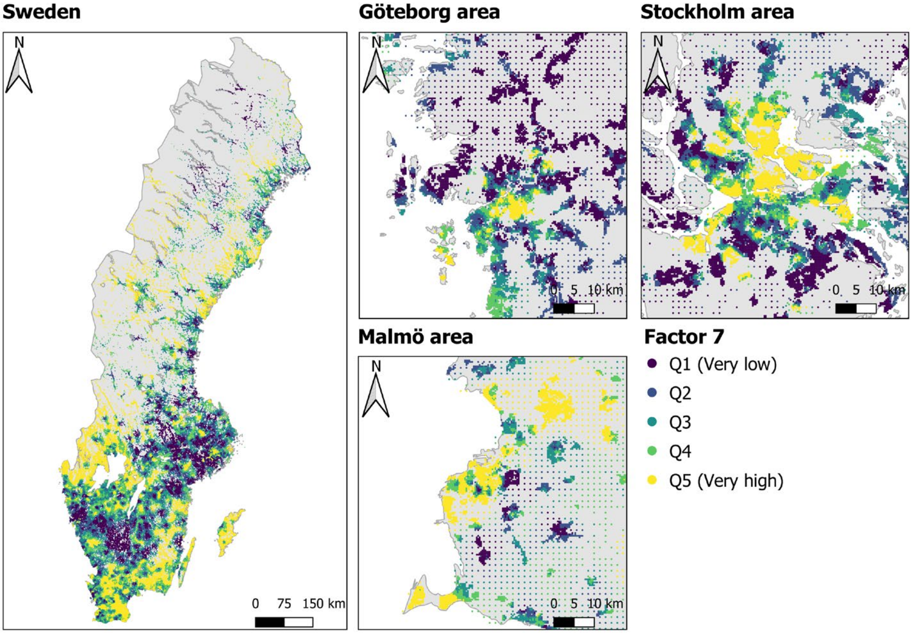
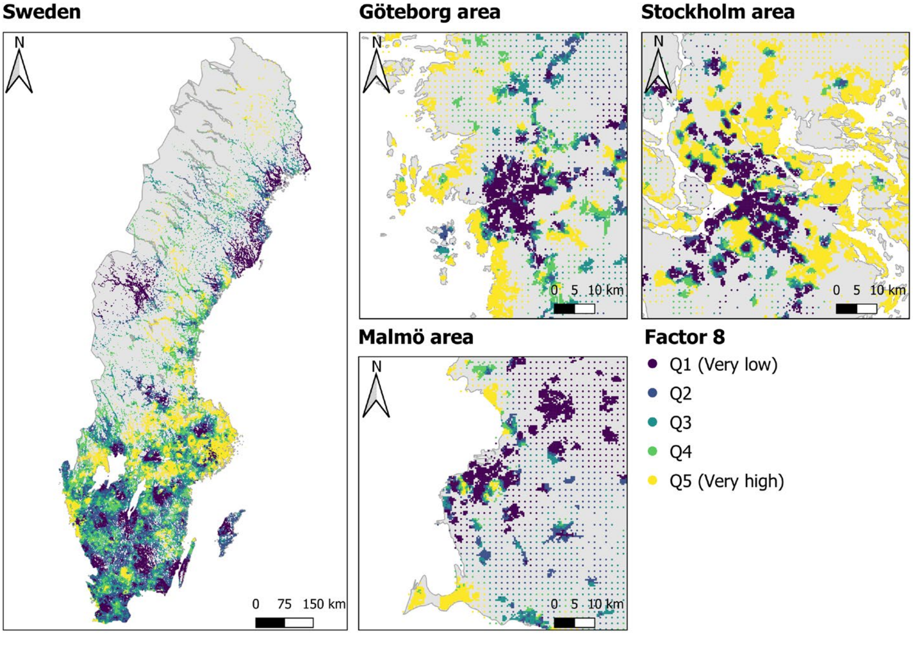


Figure 16 Large scale high income, geographical distribution



As a result of the factor analysis every populated grid cell in Sweden can be assigned 8 factor scores. There are 213,663 such grid cells or residential areas in our data. Together, these factor scores can be seen as determining the position of different geographical location in a socio-demographic contextual space. Geographic context in this representation does not only consist of the demographic composition of a residential area. Instead, context is a composite feature that includes characteristics of both the closest environment and of more extended contexts.

Factors can be used to classify residential locations into cluster typologies that capture important dimensions in the spatial variation of geographical context. To develop a typology, we use hierarchical clustering. This method produces a nested classification of residential locations, beginning from a stage where each observation forms its own single-member cluster and ending at a stage where all observations are in one single group. Several choices exist for the *linkage method*, which specifies what should be compared between groups that contain more than one observation. We chose *Ward's minimum variance method* which minimizes the total within-cluster variance (Ward 1963). This means that at each step the pair of clusters with minimum between-cluster distance are merged. Using this method, we classified residential locations into cluster-solutions ranging in size from 2 to 15 clusters. To determine the optimal number of clusters we followed the average silhouette method, inspected average silhouette plots and cluster specific silhouette plots for solutions with different number of clusters (Kaufman & Rousseeuw 2005). For calculating silhouettes coefficients we used clues R package (Chang et al. 2010). This approach is preferred for large datasets since many clustering statistics are unavailable for datasets as big as ours. Silhouette approach measures the quality of a clustering by quantifying how well each observation lies within its own cluster as opposed to the neighboring clusters. Using these indicators, our concern was to have enough clusters to capture underlying variability in the data, while maintaining sufficiently high quality of each individual cluster. In the end, we decided to break one cluster from 10-cluster solution into two further clusters from 14-cluster solution where one subcluster represents core of Urban diverse (RT_DIV core) areas and another subcluster, with lower quality of clustering based on low average silhouette value, is a buffer zone for Urban diverse cluster (RT_DIV buffer)[1]. Our typology was used for survey carried out by Statistics Sweden. For the survey we decided to exclude Rural border area, because it has poor quality of clustering and is home to relatively small number of people. We have also excluded Rural diverse buffer subcluster as it is the source of the lower overall clustering quality for Urban diverse cluster. Figure 17 to Figure 23 show the average value for the 7 socio-demographic indicators for different k-values by cluster in 10-cluster solution. Names given to clusters are based on their characteristics as well as

[1] More details can be found in the appendix.

geographical distribution of residential areas in each cluster. This is important as some of the patterns that are observed predominantly in rural areas can also be seen in larger cities[2].

[2] One example is that *Rural town diverse* (RT_DIV) cluster can also be found in Malmö and partly in Goteborg. This does not mean that these cities are classified as rural areas, it merely signals that the type of diverse areas that we see there is common in smaller towns and quite different from diversity seen for *Urban diverse* (U_DIV) cluster, which is the dominant type in Stockholm.

3. Sweden divided into ten clusters

In this chapter, we present the multiscale cluster typology of residential areas in Sweden. Table 2 presents the overview of 10 clusters we identified. We include information whether a given cluster was used for sampling by Statistics Sweden in a survey of residents living in different type of neighborhood context. Because of low clustering statistics or small number of residents we exclude one cluster (R_BOR) and narrowed sampling in another (U_DIV) to a subcluster consisting of core areas. Five clusters are identified as predominantly rural and five as predominantly urban, although there are some exceptions. This classification provides new findings. First, diversity and recent non-Western immigration is not limited to large cities and urban areas more generally. Residential areas classified as small-scale *Rural town diverse* (RT_DIV) are spread evenly across small and medium sized towns. These changes are recent. Secondly, and high income have the highest variability between cluster types. Figure 17 shows that for tertiary education the range for k=200 is bigger than the range for other indicators (0,36). The second indicator that varies a lot between clusters is the share of residents with income in the top decile, with a range equal to 0,26 for k=200. In contrast to education, this large range for high income is largely due to *Urban elite* (U_ELI) cluster having much higher share of high-income earners than other clusters.

Table 2 Cluster names and description

Name	Cluster	Key characteristics	Used in survey
Rural town diversity	RT_DIV	Small scale migration	Yes
Rural town adjacent	RT_ADJ	Adjacent to social assistance	Yes
Rural town working-class	RT_WC	Employed with low income, EU migrants	Yes
Rural homogenous	R_HOM	Few migrants	Yes
Rural border area	R_BOR	Low registered income, EU migrants	No
Urban diverse	U_DIV	Large scale migration	Core subcluster
Urban adjacent	U_ADJ	High contrast over scales	Yes
Urban homogenous	U_HOM	Medium academic with high income	Yes
Urban academic	U_ACA	Academic with medium income	Yes
Urban elite	U_ELI	Academic with high income	Yes

Figure 17 Share tertiary educated by cluster

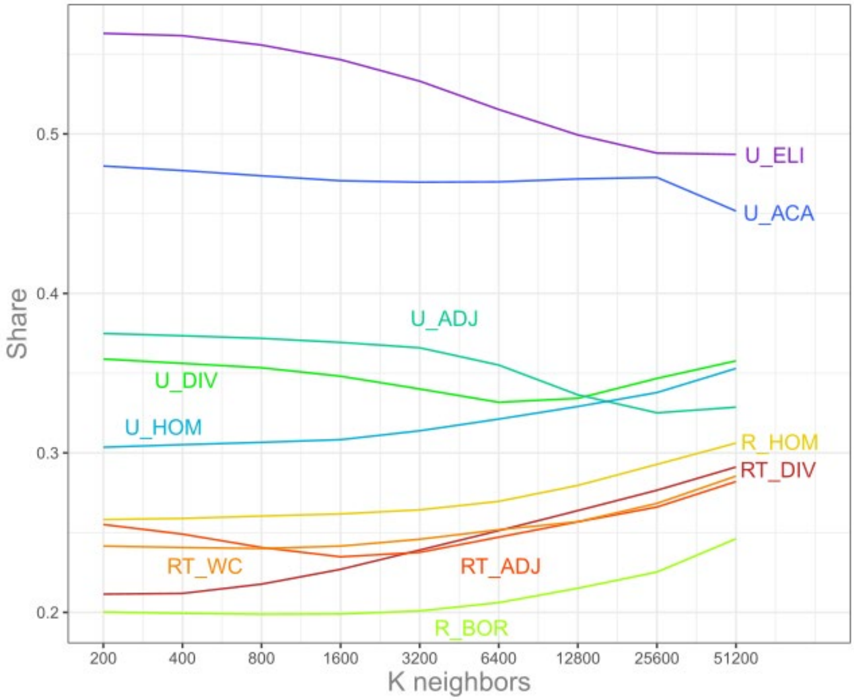


Figure 18 Share employed by cluster

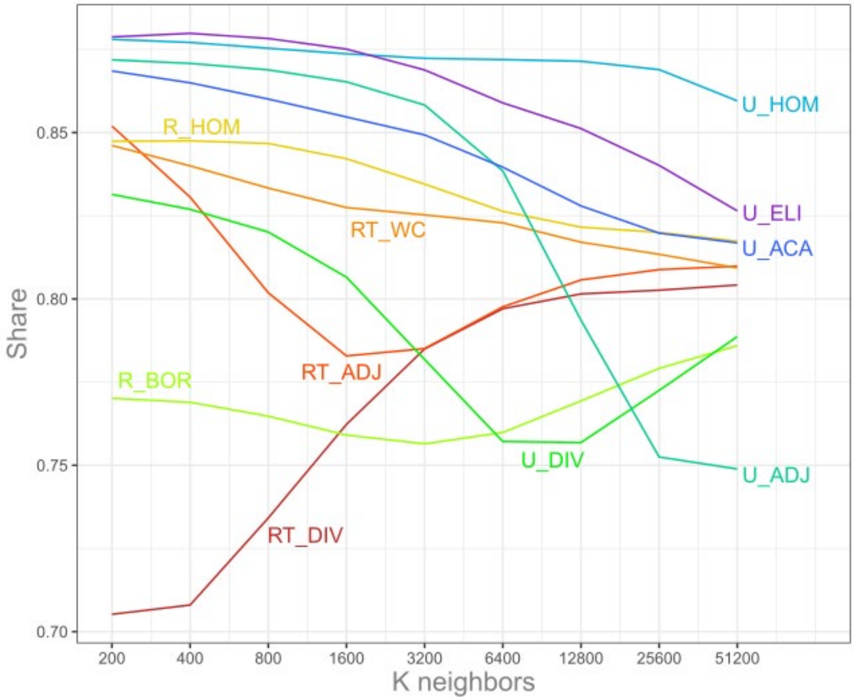


Figure 19 Share high income by cluster

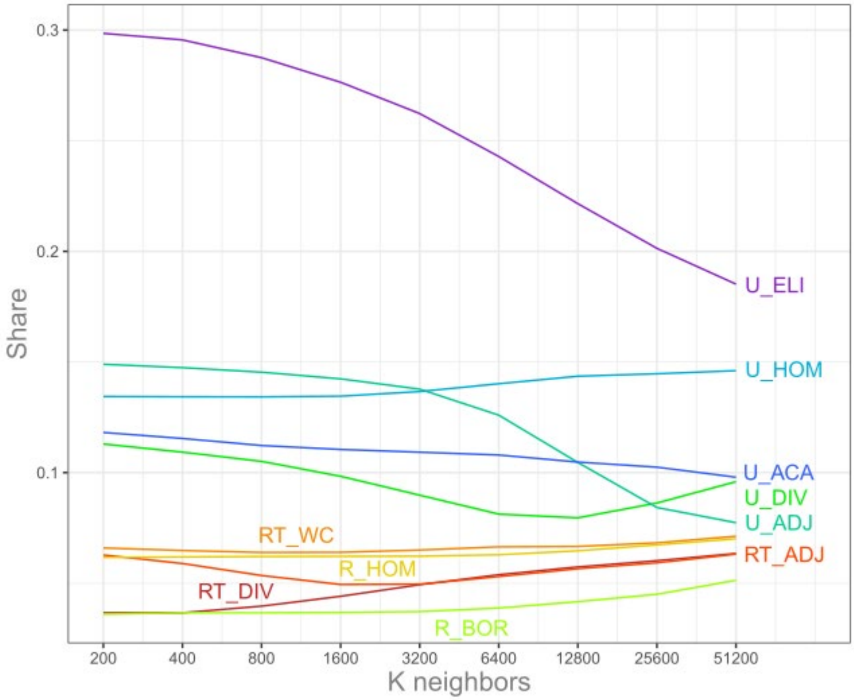


Figure 20 Share social assistance by cluster

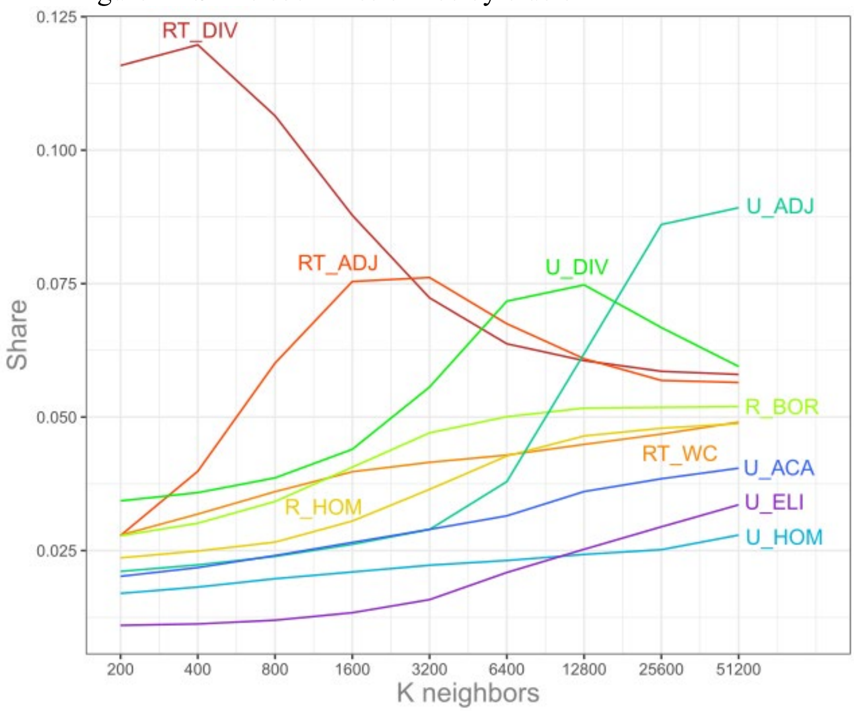


Figure 21 Share poverty by cluster

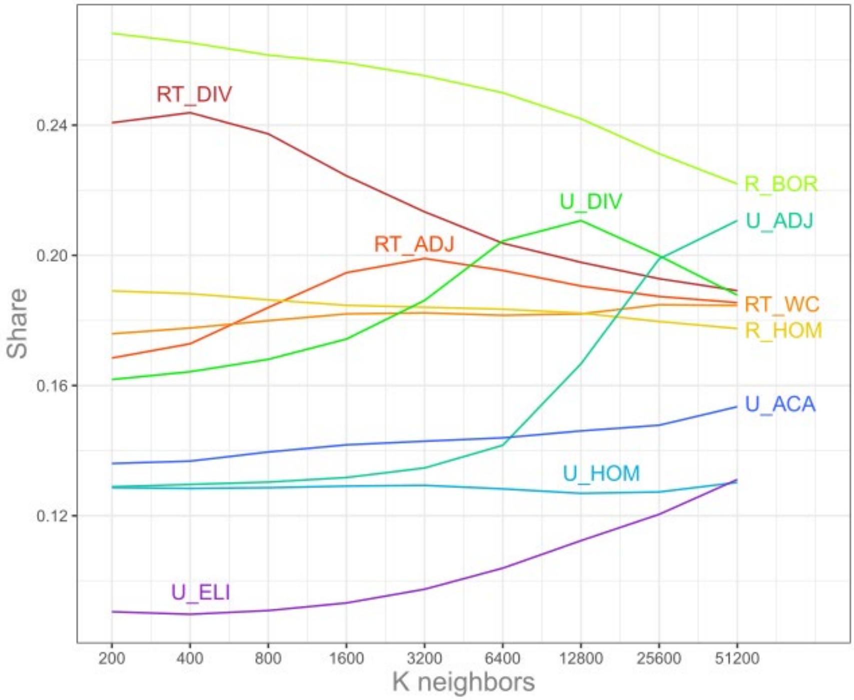


Figure 22 Share European immigrants by cluster

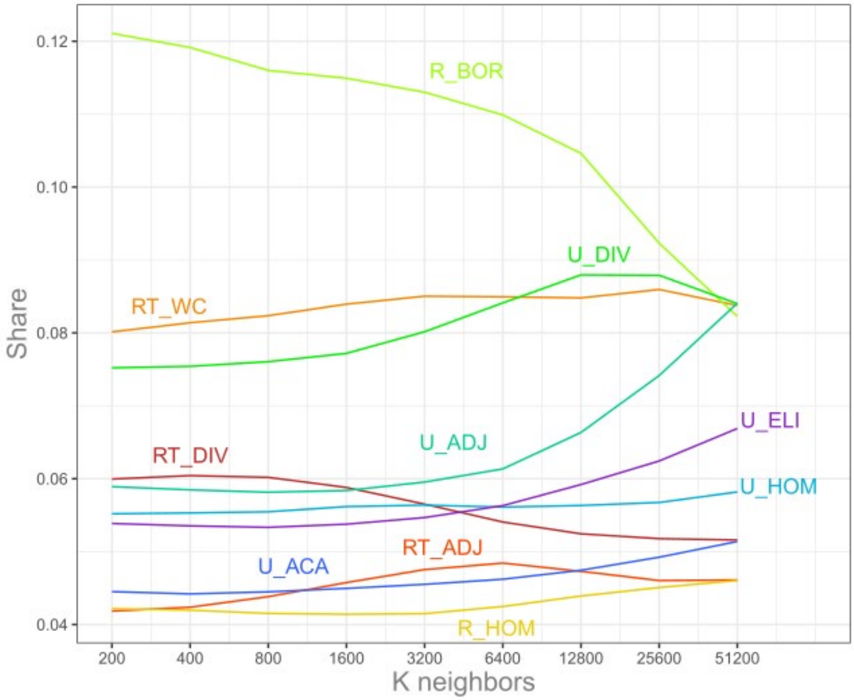
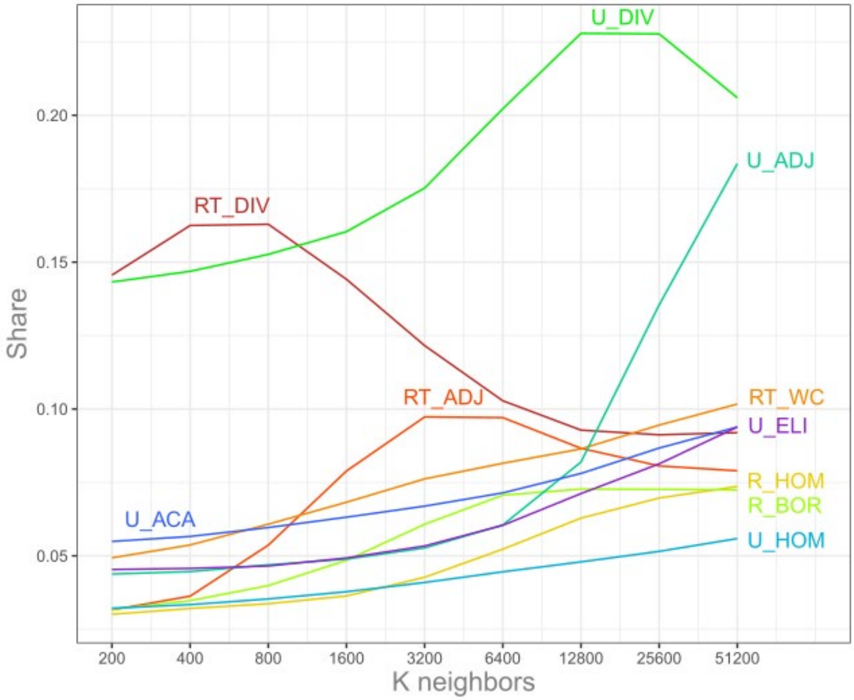


Figure 23 Share non-European immigrants by cluster



Next, for each cluster we can see which indicator shows the largest variation over scales. This says which indicator offers the most spatial contrast for a given cluster. For urban clusters these are: share non-European immigrants, high-income, employment, and tertiary education. For rural clusters, these are employment, share non-European immigrants and tertiary education (see Table 3).

Table 3 Clusters and indicators with largest deviation over scales

Cluster	Indicator	Largest difference
RT_DIV	Employed	0.10
RT_ADJ	Employed	0.07
RT_WC	non-EU	0.05
R_HOM	Tertiary	0.05
R_BOR	Tertiary	0.05
U_DIV	non-EU	0.08
U_ADJ	non-EU	0.14
U_HOM	Tertiary	0.05
U_ACA	Employed	0.05
U_ELI	High income	0.11

Education and high income are the two key variables distinguishing urban clusters from rural clusters. This means that for these indicators, the lines for urban clusters never intersect with lines for rural clusters. Education is also one indicator for which the gap between clusters remains the largest at k=51,200, the largest scale. This means that both at small and large scales, education remains the main indicator distinguishing between clusters. In other words, education is the most polarizing of all indicators we look at.

Urban clusters have also more contrast over scales. On way to assess the most contrasting clusters is to see for each indicator for which clusters we observe the largest differences over scales. Such calculation presented in Table 4 below. Five out of seven largest differences are observed for urban clusters, meaning that urban landscape offers the starkest contrasts over scales. For employment, non-European immigrants, poverty, and social assistance this is Urban adjacent cluster and for high-income it is Urban elite.

Table 4 Indicators and clusters with largest deviation over scales

Indicator	Cluster	Largest difference
Tertiary education	RT_DIV	0.08
High income	U_ELI	0.11
Employment	U_ADJ	0.12
Social assistance	U_ADJ	0.07
At risk of poverty	U_ADJ	0.08
non-European immigrants	U_ADJ	0.14
European immigrants	R_BOR	0.04

Finally, two clusters stand out with respect to immigration: *Rural town diverse* (RT_DIV) and *Urban diverse* (U_DIV), the former stands for the high share of non-European immigrants at smaller scales (peaks at k=800) and the later at large scales (peaks at k=12,800). The main difference between them is that *Urban diverse cluster* (U_DIV) is more affluent and also older. Below we present a more detailed description of clusters. To assess geographical isolation of clusters we calculated insolation and exposure indexes. This shows which types of cluster co-occur in DeSO neighborhoods and which rarely or never share neighborhoods. Full results are shown in the appendix.

Table 5 Immigrants and their date of arrival, by cluster

	All immigrants			**Non-Western immigrants**		
Cluster	Share	Mean	Median	Share	Mean	Median
RT_DIV	0.31	2003	2009	0.24	2007	2010
RT_ADJ	0.09	1995	2000	0.05	2003	2006
RT_WC	0.18	1996	2000	0.09	2003	2006
R_HOM	0. 08	1996	2003	0.04	2004	2008
R_BOR	0.18	1997	2004	0.05	2008	2013
U_DIV	0.33	1999	2003	0.24	2001	2004
U_ADJ	0.14	1994	1997	0.07	2000	2002
U_HOM	0.1	1994	1999	0.05	2001	2005
U_ACA	0.16	1998	2002	0.09	2001	2004
U_ELI	0.12	1995	2000	0.06	1999	2001

3.1 Rural town diverse (RT_DIV)

This cluster can be observed all over Sweden. Is most characteristic for small and medium sized towns, although it can also be observed in Malmö and the suburbs of Goteborg. For small scales, this cluster has the highest shares of non-European immigrants. This is coupled with higher-than-average uptake of social assistance (especially at smaller scales) and relatively low employment levels. RT_DIV is home to 14% of Swedish residents. The mean age is 43.77, the highest among all clusters. This is surprising because non-European immigration tends to go hand in hand with a more youthful population. Table 5 suggests that immigration in areas belonging to RT_DIV is more recent than in other cluster types. One likely explanation is that many areas belonging to this cluster had been ageing and depopulating until a recent influx of non-European immigration made the population more youthful and numerous again.

Together with *Urban diverse* (U_DIV), *Rural town diverse* (RT_DIV) cluster is one of the two clusters with the most diverse population. For RT_DIV this is especially visible for the share of non-European immigrants, which reaches 0.16 for k-value=400. The levels of employment, social assistance and non-European immigrants vary a lot over scales. This suggests that RT_DIV tends to represent small diverse enclaves where socio-economic conditions are different from conditions in the larger neighboring context. For employment, the level starts at 0.70 for k-value=200 and reaches 0.80 for k=51,200. For tertiary education, for k-value=200 it has the second lowest value (0.21), just after *Rural border* (R_BOR) cluster. Cluster RT_DIV has the highest share of social assistance recipients for k-value=200 (0.11) only at k-value=3,200 it is overtaken by *Rural town adjacent* cluster (RT_ADJ), which is a cluster surrounding RT_DIV. Economic disadvantage is also visible in the high share of residents who are at risk of poverty and low share of those who are in top income decile. For k-value=200, nearly a quarter of residents are at risk of poverty while the share with high income research 0.05 only after k-value=3,200.

RT_DIV is evenly spread across Sweden with residential areas belonging to this cluster present in 235 out of 290 municipalities. The highest shares of residents in *Rural town diverse* (RT_DIV) cluster live in Östra Göinge, Bjurholm, Perstorp and Högsby. In these municipalities RT_DIV residential areas are concentrated in small towns. There are eleven municipalities where over half of the population lives in RT_DIV areas. There are no visible patterns of geographical concentration and the only two municipalities with the majority of RT_DIV residential areas which are adjacent are Åsele and Bjurholm, both in the Västerbotten region. With low value for isolation index (0.54) RT_DIV residential areas tend to belong to same neighborhoods as residential areas belonging to *Rural homogenous* (R_HOM) and *Rural town adjacent* (R_ADJ) clusters.

Figure 24 Rural town diverse cluster, geographical distribution

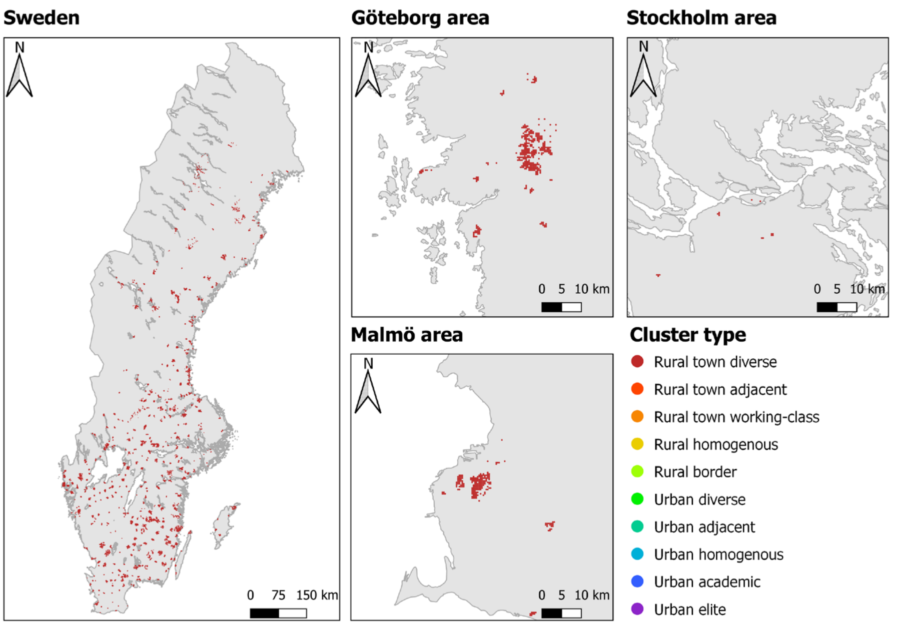
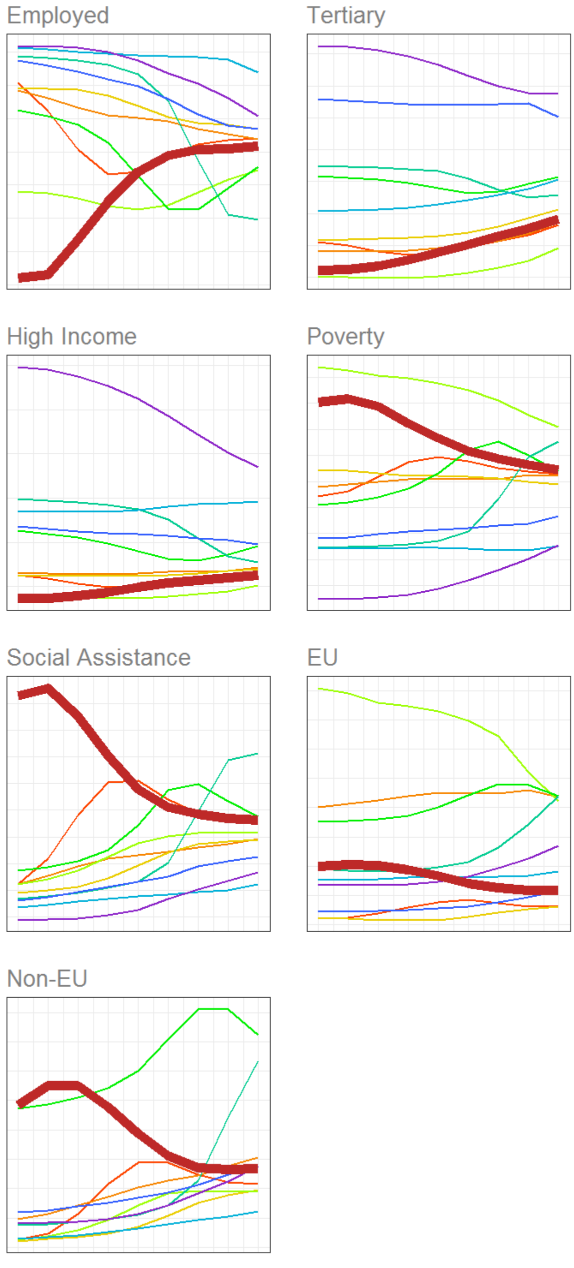


Figure 25 Rural town diverse cluster, by indicator



3.2 Rural town adjacent (RT_ADJ)

RJ_ADJ cluster signals adjacency to low employment and high uptake of social assistance. For this cluster economic disadvantage is linked to diversity. We can see that the peak in the share of non-European immigrants coincides with the peak for social assistance and low employment (for k-value=3,200). Following this peak, the share of non-European immigrants and residents receiving social assistance are declining. This suggests that residential areas belonging to RT_ADJ are bordering small enclaves with non-European immigration (and more recent wave of non-European immigrants who have had less time to integrate into the Swedish labor market), rather than larger scale segregated areas, which are common especially in the Stockholm region. *Rural town adjacent* cluster is home to 9% of Swedish residents. The average age in this cluster is 41.29 years, older than the Swedish population in general.

The share in employment starts at similar level as *Rural homogenous* (R_HOM) and *Rural town working-class* (RT_WC) clusters but drops quickly from the initial value of 0.83 to 0.78 for k-value=1,600 and then bounces back to 0.81 at the largest scale. Such a U-shape pattern suggests that RT_ADJ residential areas are often bordered by another cluster with lower employment. For tertiary education, RT_ADJ behaves similarly to *Rural town working-class* (RT_WC) cluster, starting at 0.26 for k-value=200 and ending at 0.28 for k-value=51,200. An interesting pattern can be seen for the share of residents who receive social assistance. For k-value=200 it looks similar to RT_WC, R_BOR or R_HOM but then its level increases rapidly and for k-value=3,200 it reaches a peak at 0.08, the highest among all clusters. For poverty, the share of residents with disposable income below 60 percent of the Swedish median is 0.17 for k-value=200 and 0.18 for k-value=51,200 but we also observe a moderate increase for scales between k=800 and k=12,800. For high income, we see a reversed pattern, with the line falling close to R_HOM and RT_WC but with a small decline for middle-range scales. For diversity, the line for the share of non-European immigrants follows an inverted U-shape with a peak at 0.10 for k-value=3,200. The share of European immigrants is low and does not vary much over scales, remaining below 0.05.

RT_ADJ residential areas can be seen in 205 out of 290 municipalities. Residential areas belonging to *Rural town adjacent* (RT_ADJ) cluster are evenly distributed across the country and tend to be bordered by *Rural town diverse* (RT_DIV) cluster areas. Municipalities with the highest share of this cluster type are: Hällefors, Hudiksvalls, Överkalix, Tibro and Munkfors. Hällefors has the highest share, in this municipality 72% of residents live in RT_ADJ residential areas and for other municipalities it is above 60%. Except for Hudiksvalls, a common feature of these municipalities is that they center around middle-sized towns, which had been experiencing a population decline or stagnation until the recent decade. Even though it is evenly spread, this cluster does not tend to be the dominant type. There are only 14 municipalities where most people live in residential areas belonging to this cluster. RT_ADJ is not as highly isolated as some other rural clusters (isolation index equals 0.56) and it falls in same neighborhoods as other rural cluster: R_HOM (0.26) and RT_DIV (0.12). This fits the notion that RT_ADJ is often a buffer cluster between clusters R_HOM and R_DIV, an intermediate zone between the Swedish countryside and small and medium size towns.

Figure 26 Rural town adjacent cluster, geographic distribution

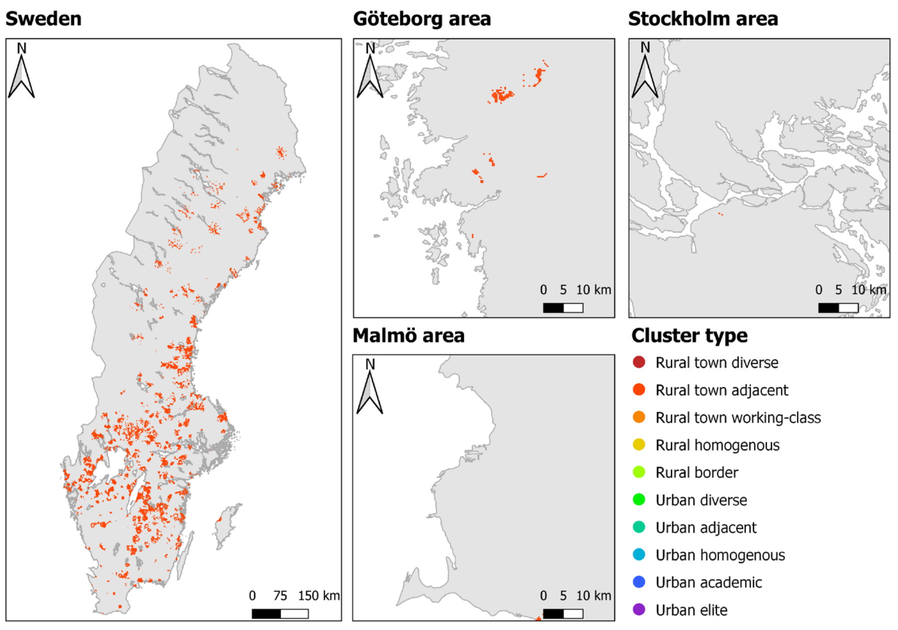
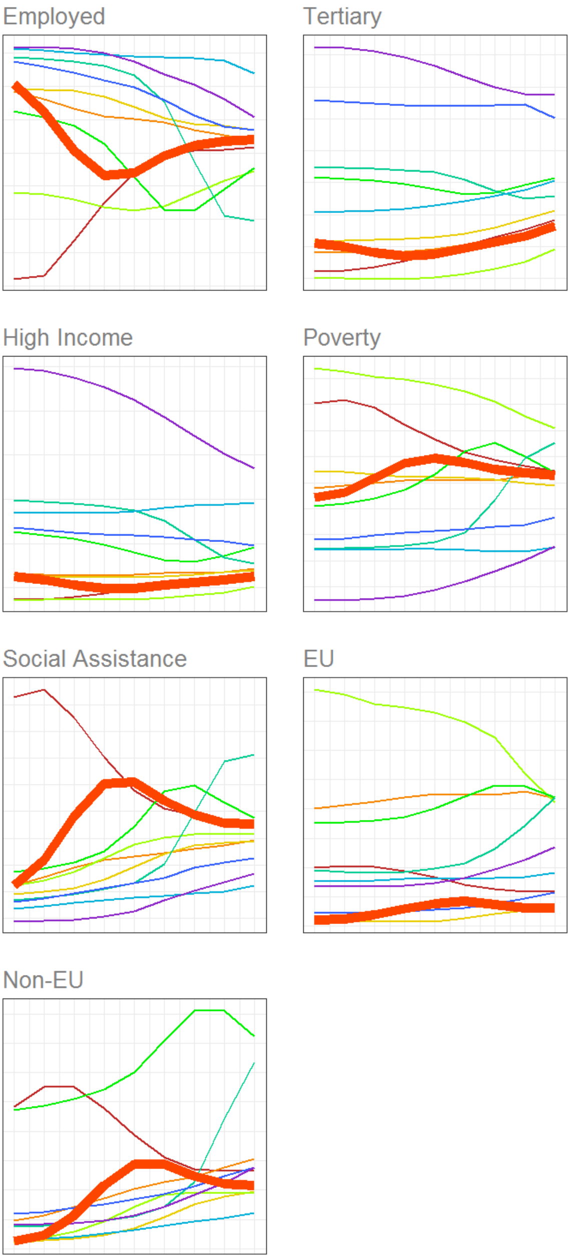


Figure 27 Rural town adjacent cluster, by indicator



3.3 Rural town working-class (RT_WC)

Residential areas belonging to *Rural town working-class* (RT_WC) cluster can be characterized as predominantly working-class areas, with residents having average employment levels and low educational attainment. Additionally, RT_WC has high levels of European immigration, especially at smaller scales (no similar pattern is observed for non-European immigration). *Rural town working-class* cluster is home to 9% of Swedish residents. Residents of this cluster are more youthful than Swedish population at large, with mean age of 39.64 years.

In many ways *Rural town working-class* (RT_WC) cluster resembles *Rural homogenous* (R_HOM) cluster but with higher levels of diversity. Starting with employment, for k-value=200, the share of residents in employment is 0.85 but the level decreases quicker than for R_HOM. For k-value=51,200 the share of residents in employment is 0.81. RT_WC has low values for the share of residents with tertiary education, starting at 0.24 for k-value=200 and staying below 0.30 at larger scales. The share of residents who receive social assistance is 0.03 at k-value=200, increasing to 0.05 for k-value=51,200, which puts RT_WC as an average cluster with respect to take-up of social assistance. For poverty and high-income indicators, the shares at different scales look similar to R_HOM. For diversity, the share of European immigrants stays around 0.08 for all scales. For k-value=200 this is the second largest value, after *Rural border area* (R_BOR), another rural cluster characterized by high shares of European immigrants in border zones. For non-European immigrants, the level starts at 0.05 at k-value=200 and increases to 0.10 for k-value=51,200.

Rural town working-class (RT_WC) cluster is present in 137 out of 290 municipalities and is often concentrated around old industrial centers. Municipalities with the highest proportion of residents living in RT_WC residential areas are: Surahammar, Vaggeryd, Gislaved, Värnamo and Gnosjö. Historically these areas have been important for Swedish industry and mining, the first two belonging into the mining district of central Sweden while the latter are a part of the Gnosjö region, known for its industrial productions and high employment rate. Only 28 municipalities have more than half of residents living in RT_WC residential areas. These municipalities are concentrated around Kronoberg and Jönkoping regions in the South, Västmanlands in central Sweden and in Norrland, in areas close to the border with Finland. RT_WC has isolation index of 0.75, higher than previous two clusters. The highest values for exposure index are for RT_ADJ (0.07) and RT_DIV (0.03).

Figure 28 Rural town working-class cluster, geographic distribution

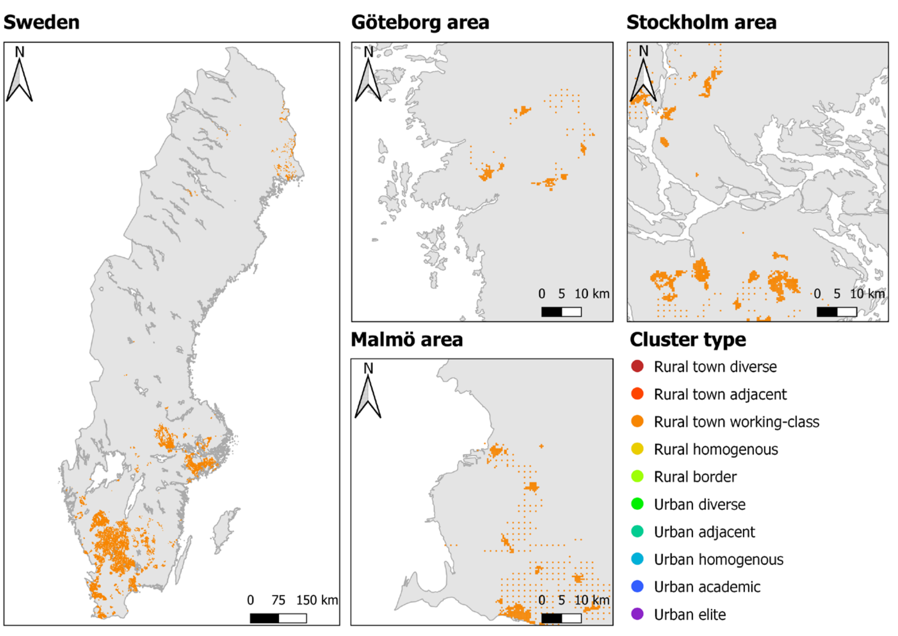
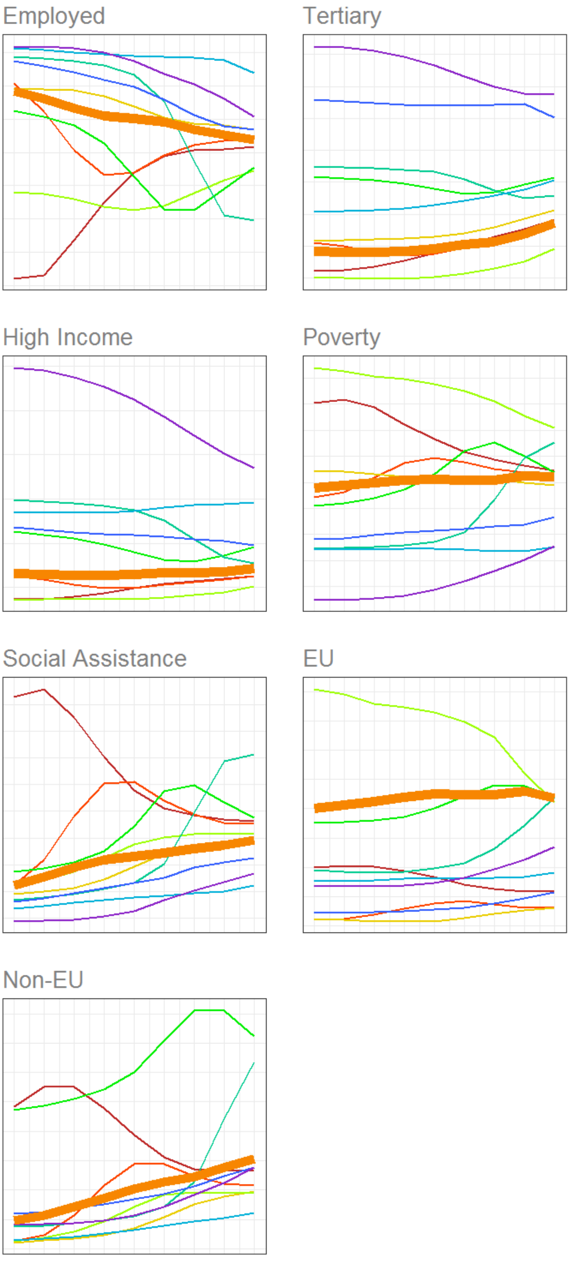


Figure 29 Rural town working class cluster, by indicator



3.4 Rural homogenous (R_HOM)

Residential areas belonging to *Rural homogenous* (R_HOM) cluster are evenly distributed across Sweden and are the most common cluster type outside of metropolitan areas. R_HOM is the average cluster with respect to most indicators, except for immigration where for all scales the share of European and non-European immigrants is lower than for other clusters. R_HOM is the most populous cluster, with 21% of Swedish residents living in residential areas belonging to this cluster type. With mean age of 42.68, residents of R_HOM are older than the Swedish average (40.69).

When compared to other rural clusters, R_HOM seems like an affluent cluster. For example, it has higher levels of employment than other rural clusters. For residential areas in this cluster, the share in employment is 0.85 at smaller k-values (200, 400 and 800) dropping to 0.82 at the largest k-value (51,200). It has relatively low values for the share of residents with tertiary education, starting from 0.25 for k-value=200 and reaching 0.31 for k-value=51,200, yet the line for the share with tertiary educated for R_HOM stays above the corresponding lines for other rural clusters (RT_DIV, RT_ADJ, RT, WC, R_BOR), which means that it has higher education attainment than other rural clusters. When it comes to social assistance, R_HOM has lower values not only than other rural clusters, but also than *Urban diverse* (U_DIV) cluster. The values for the share of residents who receive social assistance are low, starting from 0.02 at k-value=200, and reaching 0.05 for k-value=51,200. For poverty, at k-value=200 the share with disposable income below 60 percent of the Swedish median is 0.19, the third highest value (higher values at this scale can be observed for only two other rural cluster: RT_DIV and R_BOR) and poverty remains stable throughout scales. This does not necessarily mean that residents of this luster are poor. First, there are large wage disparities between rural and urban areas in Sweden. *Rural homogenous* (R_HOM) cluster covers many sparsely populated areas, where salaries are lower compared to earnings in cities and towns. Residential areas in R_HOM belong to some of the least diverse areas in Sweden. For non-European immigrants, at k-value=200 the share non-European immigrants is only 0.03, increasing to 0.08 at k-value=51,200. For the share of European immigrants, the line for R_HOM remains below all other clusters and never reaches 0.05. The value for European and non-European immigration is much below the Swedish average (in 2016 the share of foreign-born residents born outside of Europe was 0.13 and in Europe 0.08).

Rural homogenous (R_HOM) cluster is a common cluster type and it is present in 260 out of 290 municipalities. The highest share of residents in R_HOM residential areas can be found in municipalities of Båstad, Vindeln, Höör, Vansbro, Ystad and Leksand where over 90% of residents live in residential areas belonging to this cluster type. R_HOM is also the dominant cluster in terms of geographic coverage. To see this, consider that in a quarter of all municipalities, more than half of residents live in R_HOM residential areas (74 out of 290). Such municipalities are concentrated in Skåne or Blekinge region in southern Sweden, Västra Götaland municipalities (south of lake Vänern) and Jämtland or Dalarna regions in northern Sweden. The index of isolation is 0.82, which means that R_HOM is highly isolated and is rarely exposed to other cluster types in neighborhoods. When it does share neighborhoods with other cluster types it tends to be with other rural clusters (index of exposure equal to 0.07 for RT_ADJ).

Figure 30 Rural homogenous cluster, geographic distribution

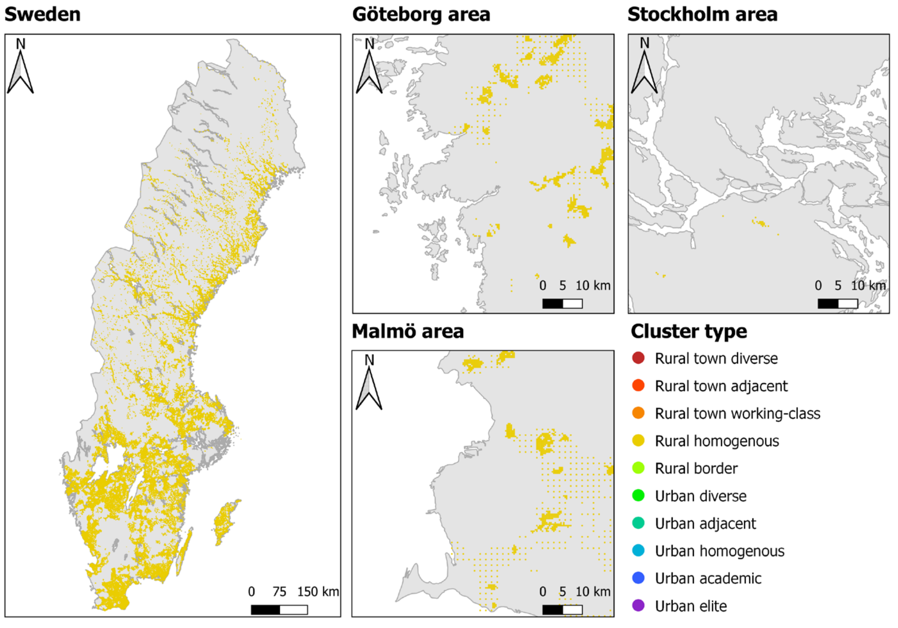
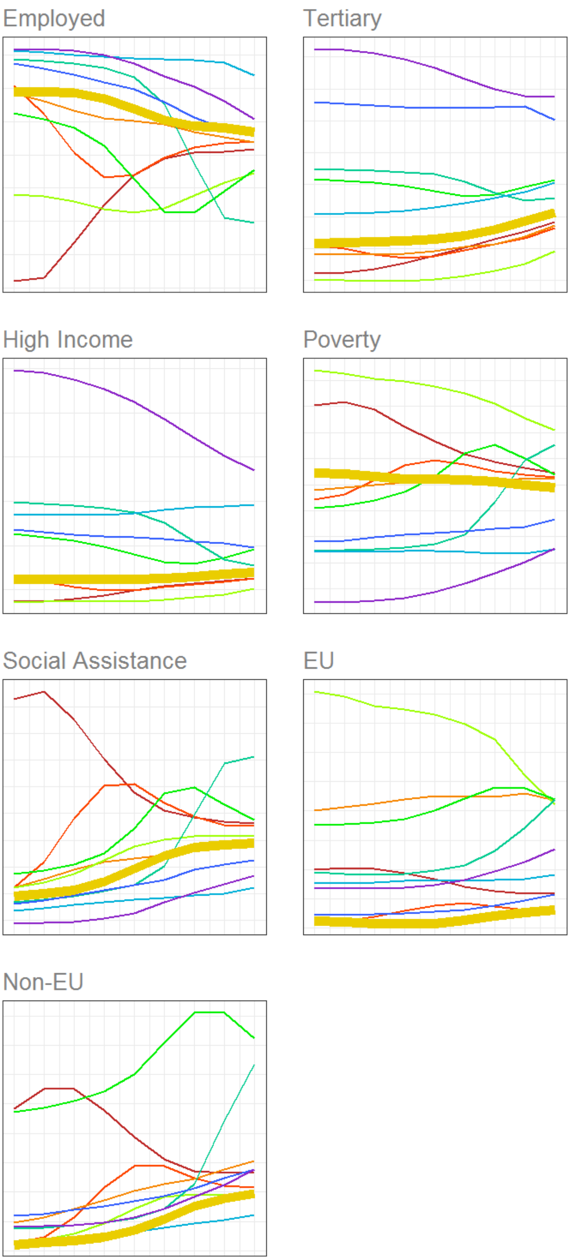


Figure 31 Rural homogenous cluster, by indicator



3.5 Rural border area (R_BOR)

The most distinctive characteristic of *Rural border areas* (R_BOR) cluster is that it has the highest shares of European immigrants at small scales, which then declines steeply at larger scales. Similarly like in case of *Rural town working-class* (RT_WC) cluster, this is not coupled with high levels of non-European immigrants. R_BOR residential areas are concentrated around Swedish border areas, in Värmlands, Östra Götlands, Skåne and Norrbottens. This cluster is the oldest (mean age is 44.52 years) and the smallest with only 1% of Swedish population living in residential areas belonging to this cluster type.

R_BOR is characterized by low employment and education levels. This cluster has the lowest share of tertiary educated residents. For k-value=200 only one in five of those aged 25-64 has tertiary education (compare this to over one in two for *Urban elite* cluster), a value which increases to one in four for k-value=51,200. Another noteworthy feature is the high level of those who are at risk of poverty – it is higher than for any other cluster. For k-value=200 the value is 0.27 and for k-value=51,200 it is 0.22. The fact that more than a fifth of working age population is in relative poverty is quite striking. Population ageing may contribute to these high levels of disadvantage. Pensioners are at a higher risk of poverty because their disposable income in lower than disposable incomes of working age individuals. Furthermore, most of R_BOR residential areas are in sparsely populated parts of Sweden where salaries are lower. Yet, R_BOR residential areas are not necessarily poor. When we look at the share of residents who receive social assistance, this cluster does not stand out as particularly disadvantaged. For social assistance, there are many clusters that overtake R_BOR. This cluster is close to border areas and it is possible that some residents commute to work and are registered as employees in other countries, for instance in Norway. R_BOR has the highest level of European immigrants, for k-value=200 the value is 0.12. Non-European immigration is relatively low. For k-value=200 the value for the share of non-European immigrants is 0.03 and it increases to 0.07 for k-value=51,200.

Rural border areas (R_BOR) cluster is present in 79 out of 290 municipalities. There are three municipalities which consists entirely of residential areas belonging to R_BOR: Strömstad, Eda and Årjäng. In these municipalities Norwegian born residents represent the largest foreign-born group. Further four municipalities add to the list of municipalities where most residents live in R_BOR residential areas: Dals-Ed, Haparanda, Örkelljunga and Klippan. The last two municipalities are perhaps the most surprising ones because they are not adjacent to any borders (although their seats are within 60 kilometers from Denmark). The index of isolation for R_BOR is 0.70, when R_BOR residential areas share neighborhoods with other clusters this tends to be R_HOM (0.15) and to a lesser degree RT_WC and RT_ADJ (0.07 and 0.06 respectively).

Figure 32 Rural border cluster, geographic distribution

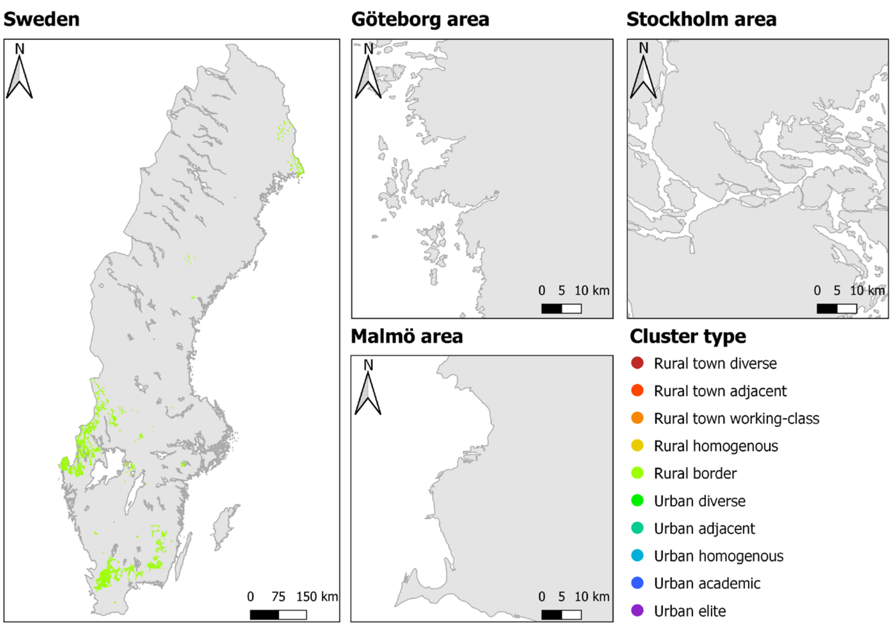
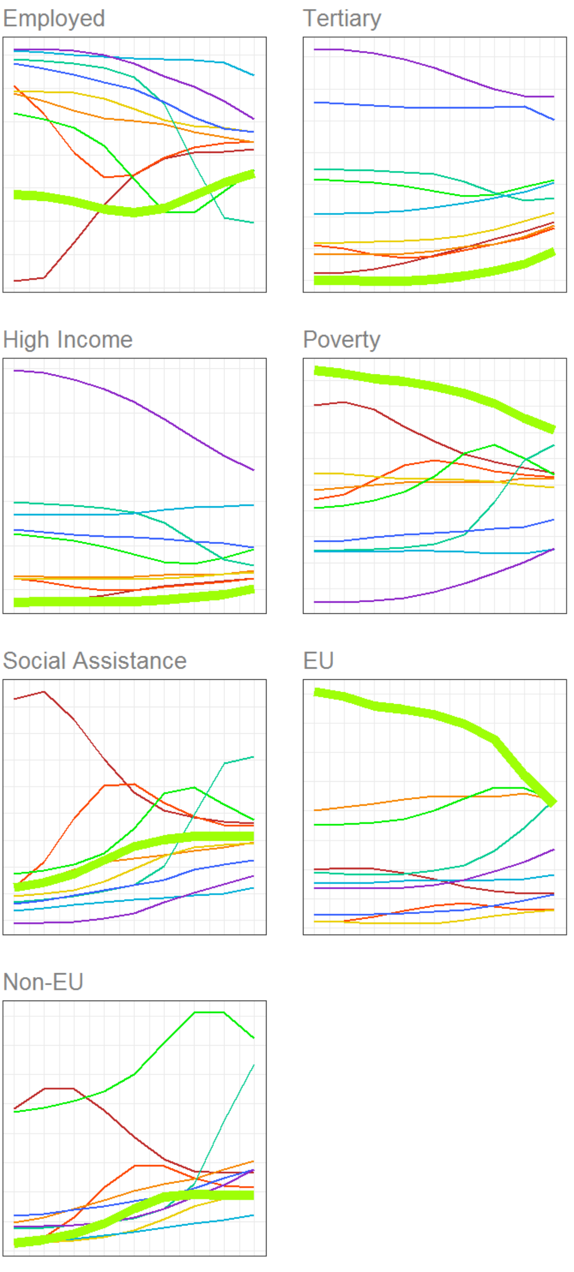


Figure 33 Rural border cluster, by indicator



3.6 Urban diverse (U_DIV)

In contrast with *Rural town diverse* (RT_DIV) cluster, *Urban diverse* (U_DIV) cluster is concentrated in and around cities. It is most common in Stockholm region and Stockholm municipality where is the third most common cluster type. U_DIV has the highest share of non-European immigrants (increasing at larger scales) and one of the highest shares of European immigrants. With the mean age for residents is 38.01, it is the youngest of all clusters. *Urban diverse* cluster is home to 11% of Swedish population.

Just as *Rural town diverse*, *Urban diverse* cluster has high levels of economic deprivation, but for U_DIV deprivation is most visible at middle scales between k-value=6,400 and k-value=12,800. This pattern arises because U_DIV picks up the macro scale of urban segregation, where individual grid cells can be on average better off than their larger surroundings. This is what distinguishes it from RT_DIV, which picks up mostly micro level segregation. It is also possible that U_DIV consists of two smaller subclusters. Indeed, after inspecting this cluster closer, we decided it could be broken into two subclusters. One represents the core and the other a buffer zone around these core areas. Plots for sub-groups of U_DIV are available in the Appendix.

When we examine Figure 35 and the corresponding figures for the subclusters in the Appendix, we see that even though there is a clear pattern of economic deprivation, U_DIV does not stand out as a cluster that has low levels of tertiary education. In fact, at all scales U_DIV has higher shares of tertiary educated residents than *Urban homogenous* (U_HOM), the affluent rural cluster type. Large dips were observed for employment (the lowest value is 0.77 for k-value=25,600), risk of poverty (0.21 for k-value=12,800) and social assistance (0.07 for k-value=12,800). The values for high income and education, while indicating deprivation, do not change so much over scales. *Urban diverse* (U_DIV) is the most diverse cluster, especially for non-European immigration. For k-value=12800 and k-value=25,600 the share of residents born in non-European county is 0.23. When we look at the subclusters, the core subcluster has the highest value for the share of European immigration equal to 0.12 for k-value=6,400. Additionally, the value for the share of non-European immigrants in the core area reaches 0.38 for k-value=3,200.

U_DIV residential areas are present in 49 out of 290 municipalities. It is the most southern cluster, with the most northbound residential areas in Gävle. The highest share of residential areas belonging to this cluster can be found in Sigtuna municipality, however in this case more than half of Sigtuna's U_DIV residential areas belong to a buffer zone subcluster. For a high concentration of the core subcluster, the only municipality with over half of residential areas belong to this category is Botkyrka (0.53). There are fifteen municipalities with the core subcluster residential areas, of which thirteen are in the suburbs of Stockholm, one is in Goteborg and one in Borås.

One interesting feature of urban deprivation is how it differs between the three largest cities. To examine this, we overlayed boundaries of police denominated *särskilt utsatta områden*, Swedish especially vulnerable areas, with distribution of diverse clusters. In the Stockholm region there are six especially vulnerable areas, and all consist exclusively of *Urban diverse* (U_DIV) core subcluster. In Goteborg municipality there are also six especially vulnerable areas, yet except for Biskopsgården, they are all *Rural town diversity* (R_DIV) residential areas. In Malmö municipality all three especially vulnerable areas are in *Rural town diversity* (R_DIV) areas. This does not mean that we classify parts of Malmö or Goteborg as rural areas, but that the patterns found there resemble patterns often found in small and mid-sized towns and cities. One reason could be the different patterns of segregation in these cities, where in Stockholm we see mostly macro segregation while in Goteborg and Malmö a mixture of macro and micro patterns. Another interpretation has to do with socio-economic deprivation, RT_DIV areas in Goteborg and Malmö are more deprived than U_DIV.

Urban diverse (U_DIV) is highly isolated. Its index of exposure to itself is equal to 0.76 (compare it to 0.56 for RT_DIV), when it shares neighborhoods with other cluster types these tend to be RT_WC (0.06) or R_HOM (0.05).

Figure 34 Urban diverse cluster, geographic distribution

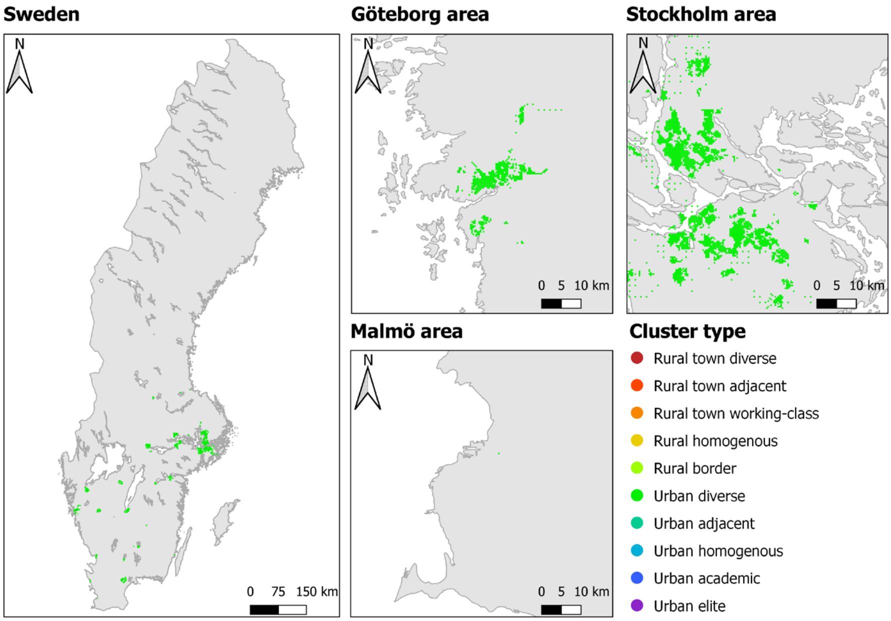
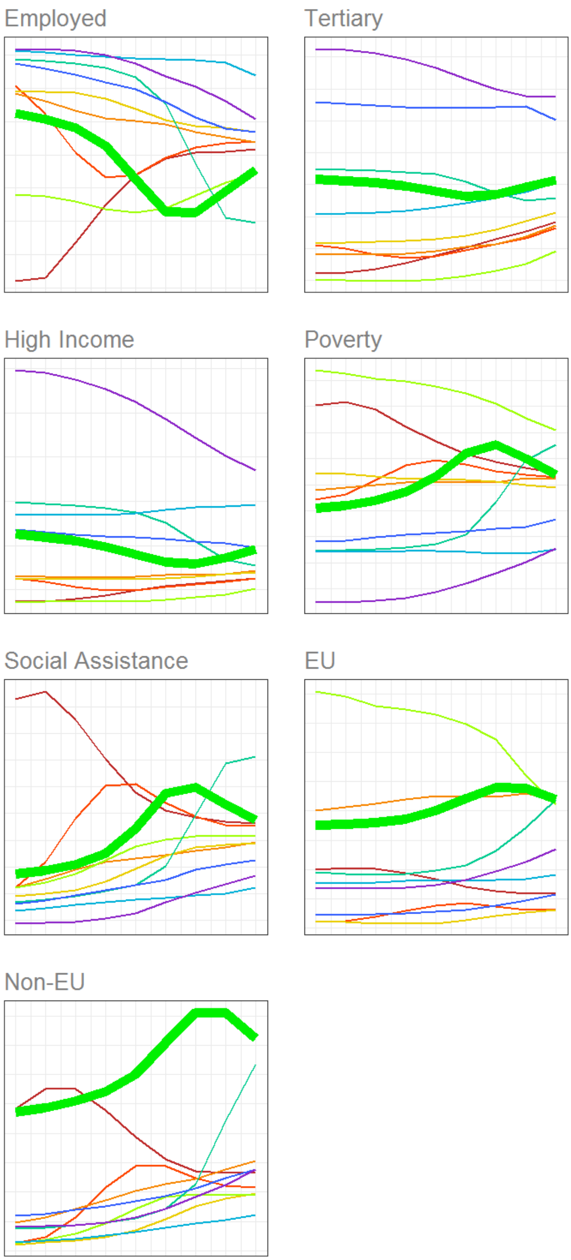


Figure 35 Urban diverse cluster, by indicator



3.7 Urban adjacent (U_ADJ)

Urban adjacent (U_ADJ) cluster is characterized having its share of immigrants starting low at small scales but increasing quickly for lager scales. This cluster is similar to two other clusters. First, it is like *Rural town adjacent* (RT_ADJ), in that it is a buffer zone for a diverse cluster. In case of RT_ADJ this buffer zone referred to *Rural town diverse* (RT_DIV) cluster and in case of U_ADJ it refers to *Urban diverse* (U_DIV). Second, U_ADJ shares some similarities with *Urban homogenous* (U_HOM), the main difference is that residential areas in U_ADJ are adjacent to urban diversity while U_HOM is homogenous and distinctly isolated from diversity. This cluster is one of the smaller clusters, with less than 4% of Swedish populating living in residential areas belonging to U_ADJ. The mean age of residents is 40.80 years, which is close to the average age of the Swedish population.

This cluster is characterized by high contrast over scales. At small scales, it looks like other middle of the range urban clusters, but it displays a distinct behavior after k-value=6,400 when it switches from relatively affluent cluster to relatively deprived one. Employment level is 0.87 for k-value=400, one of the highest values, but at k-value=51,200 it drops to 0.75, the lowest value at this scale. For education, this cluster starts as the third highest value with 0.37 having tertiary education but it drops to 0.33 at k=51,200, the lowest value for any urban clusters. For k-value=200 the share of those with income in the top income decile is 0.15, which makes U_ADJ the second most affluent cluster (after *Urban elite*). This value drops after k-value=6,400 and reaches 0.08 at k=51,000. For poverty risk and social assistance, U_ADJ starts as one of the most affluent clusters but for both indicators the values rise sharply at larger scales. This shows that U_ADJ residential areas are proximate to another, more deprived cluster. Further examination reveals that U_ADJ tends to act as a buffer zone between areas of urban deprivation and urban or rural affluence. The deprived zone that is adjacent to U_ADJ is also more diverse, especially in terms of presence of non-European immigration.

U_ADJ residential areas are present in 35 out of 290 municipalities. There is one municipality, Nykvarn, with 99% of residential areas belong to U_ADJ. This municipality is located between the municipality of Södertälje, diverse suburbs of Stockholm on the East and Gnesta and Strängnäs, two predominantly rural municipalities in Södermalm region on the West. Two further municipalities consist mostly of U_ADJ residential areas: Söderköping and Ekerö. With the most northern point in Gävle, U_ADJ is another cluster (next to U_DIV) which is heavily concentrated in southern Sweden. The index of isolation is equal to 0.77. Clusters which tend to belong to same neighborhoods with *Urban adjacent* cluster are R_HOM (0.07), RT_WC (0.04) and U_DIV (0.06).

Figure 36 Urban adjacent cluster, geographic distribution

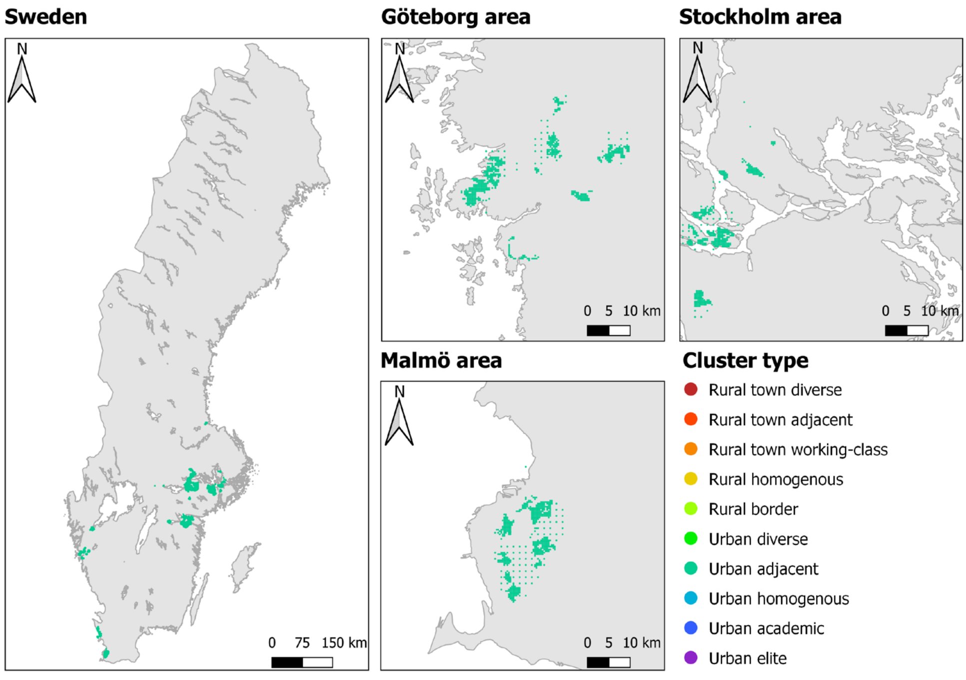
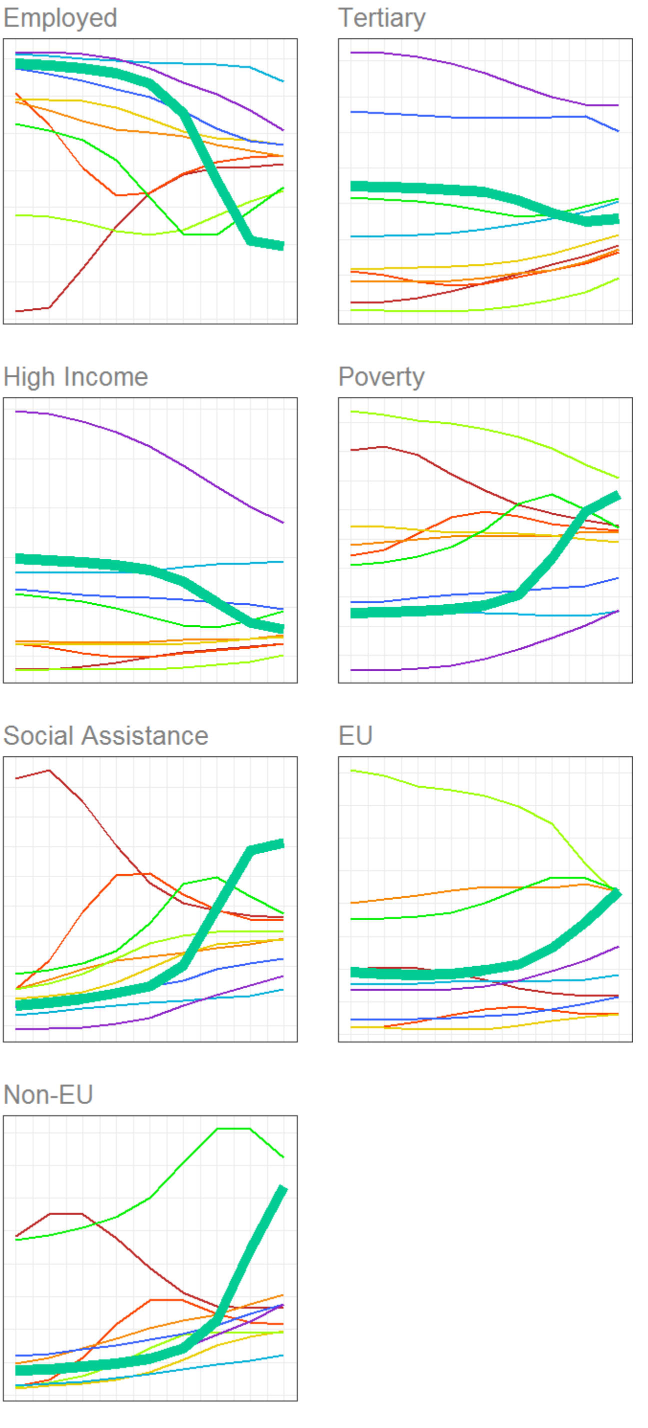


Figure 37 Urban adjacent cluster, by indicator



3.8 Urban homogenous (U_HOM)

Urban homogenous (U_HOM) cluster is characterized by high levels of employment, starting at the same level as *Urban elite* (U_ELI) cluster which do not decline over scales. U_HOM also has low shares of non-European immigrants, the second lowest share of residents at risk of poverty and residents who receive social assistance. Residential areas in this cluster are home to 6% of Swedish population. The average age of its residents is 40.76 years, close to the Swedish average.

This cluster is characterized by a strong labor market attachment. The share in employment starts at 0.88 at small k-values and declines only slightly to 0.86 for k-value=51,200. From k-value=3200 the share in employment overtakes that in *Urban elite* (U_ELI) cluster. Related to employment is the second lowest take-up of social assistance and low share of residents at risk of poverty. Yet, these indicators of economic security are not matched with high income. For k-value=200 the share of residents with income in the top income decile is 0.13, the third highest value, which increases for larger scales reaching 0.15 for k-value=51,200. This means that on average, proximate areas that are further away are more economically affluent than proximate areas which are close. Since income is strongly correlated with education, lower share of top income earners could be explained by lower educational attainment of residents in this cluster. U_HOM stands out as having the lowest share of tertiary educated residents among all five urban clusters, lower than *Urban diverse* (U_DIV) cluster. For small scale (k-value=200) the share with tertiary education is 0.30, a value which goes up to 0.35 for k-value=51,200. This increase is observed because U_DIV is an urban cluster and at larger scales the values are influenced by proximity to other urban clusters (especially U_ELI and U_ACA). U_HOM can be described as the only urban cluster, which is clearly isolated from diverse areas, especially from non-European immigrants. The value for the share of non-European immigrants reaches 0.05 for k-value=12,800 and it ends at 0.06 for k-value=51,200, the lowest value among all clusters. At the same time, U_HOM does not stand out in terms of its share of European immigrants. It starts with a value that puts U_HOM in the middle of other cluster types and ends at roughly the same level. Note that this cluster, together with *Rural border areas* (R_BOR) are the only two clusters for which values for European immigration are higher than for non-European immigration.

Urban homogenous (U_HOM) residential areas are present in 51 out of 290 municipalities. In the Stockholm region this cluster could be called the "Skärgårds cluster" because it follows a distinctive pattern where residential areas belonging to this cluster are found in the Stockholm archipelago. In is also present on the Western coast - Tjörn, an island municipality which belongs to Götenborg archipelago consists exclusively of U_HOM residential areas. Other municipalities where U_HOM residential areas constitute at least 90% are: Öckerö, Vallentuna, Österåker, Gällivare and Kiruna. The last two are in Norrboten, the norther part of Sweden. The seats of these municipalities are two northern towns, both important mining centers. This could explain high employment which is not coupled with high educational attainment or diversity. U_HOM is the most isolated cluster with index of isolation equal to 0.88. This means that it rarely shares neighborhoods with residential areas from other cluster types. When it does, it tends to share with R_HOM cluster (0.05).

Figure 38 Urban homogenous cluster, geographic distribution

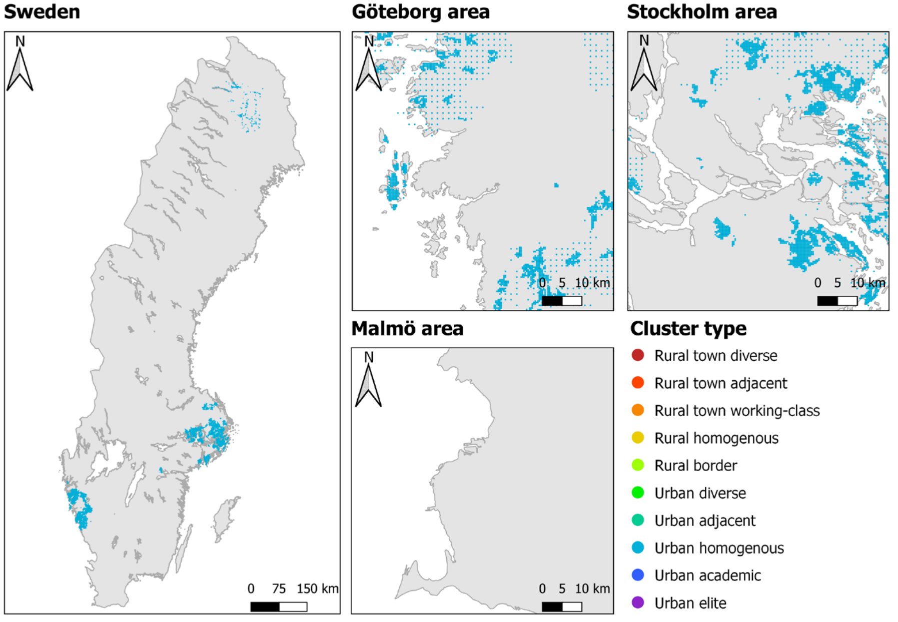
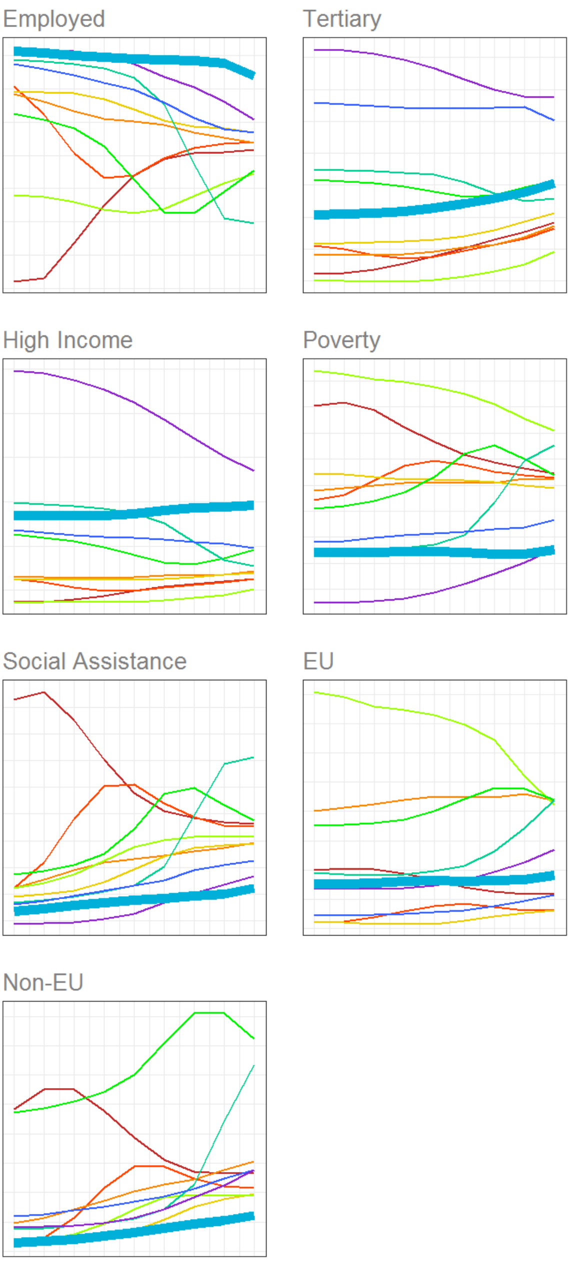


Figure 39 Urban homogenous cluster, by indicator



3.9 Urban academic (U_ACA)

Urban academic (U_ACA) cluster is similar *Urban elite* (U_ELI) cluster with respect to high educational attainment, but here education is not coupled with economic affluence represented by the share of residents with income in the top income decile. U_ACA is characterized by strong labor market attachment, low poverty rates and scant presence of immigrants. Together with U_ELI it is the most prominent cluster type for residential areas in Stockholm municipality. Approximately 17% of Swedish population lives in U_ACA residential areas and the mean age is 39.14 years.

U_ACA residential areas are a good example that economic and social affluence does not always go hand in hand. At small scales (k-value=200) U_ACA is placed as the fourth most economically affluent cluster out of five urban clusters, with only one urban cluster (*Urban diverse* U_DIV) having lower levels for employment, share of top income earners and higher levels of share at risk of poverty. Yet, U_ACA stands out in terms of education, being the second cluster (after *Urban elite* U_ELI) with the highest share of tertiary educated residents. For k-value=200, the share of tertiary educated is 0.48 and it declines to 0.45 for k-value=51,200. U_ACA is one of the least diverse clusters when it comes to European immigration, especially at small scales the share of European immigrants is low, and it remains among the lowest at larger scales. Yet, for non-European immigration, U_ACA does not stand out as a particularly homogenous cluster. In fact, for non-European immigration, at both the small and large scales it is the third most diverse cluster, although it should be noted that there is a large gap between the two most diverse clusters (RT_DIV and U_DIV) and other clusters.

Urban academic (U_ACA) cluster can be found all over Sweden, especially in proximity to middle-size university towns. Residential areas belonging to this cluster are present in 56 out of 290 municipalities. In Solna and Lund the share of U_ACA residential areas is above 90%. Other municipalities with more than half of residential areas in U_ACA are: Umeå, Hammarö, Östersund, Mörbylånga, Luleå, Habo, Linköping, Uppsala, Växjö and Mölndal. All these municipalities have a university town or are adjacent to another municipality with a university town. With index of isolation equal to 0.80, U_ACA is one of the highly isolated clusters and rarely shares neighborhoods with other cluster types. When it does, it tends to be R_HOM (0.12).

Figure 40 Urban academic cluster, geographic distribution

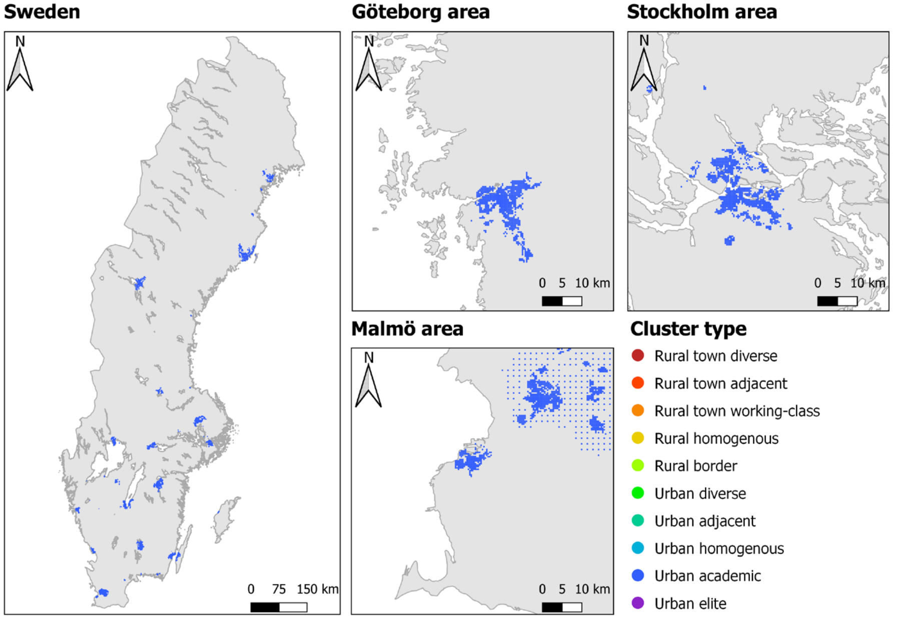
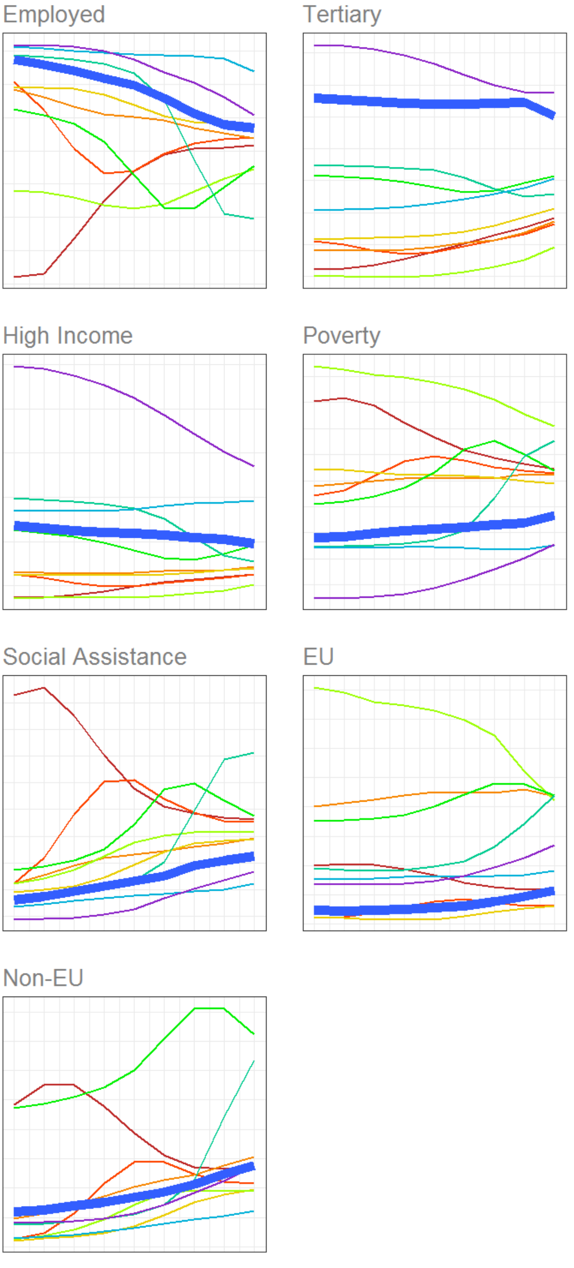


Figure 41 Urban academic cluster, by indicator



3.10 Urban elite (U_ELI)

Urban elite (U_ELI) cluster is characterized by exceptionally high educational attainment coupled with exceptionally high income. U_ELI has the highest share of high-income earners, especially at smaller scales. It also has the lowest levels of poverty and social assistance, although this increases slightly for larger scales, similar with the presence of European and non-European immigrants. This is because residential areas belonging to this cluster are concentrated in urban areas, which are also home to diverse immigrant population. Around 8% of Swedish population lives in residential areas belonging to U_ELI. This cluster is most common in cities, especially Stockholm but can also be found in proximity to some university towns, for instance around Uppsala or Linköping. The mean age is 39.05, slightly below the average age for Swedish residents.

U_ELI has a distinct profile of high socio-economic status. This cluster has the highest values of employment up to k-value=1,600 after which it is overtaken by *Urban homogenous* (U_HOM) cluster. For k-value=200 the share in employment is 0.88 and falls to 0.83 for k-value=51,200. *Urban elite* is also the most affluent cluster in terms of educational attainment. For all scales, the values for the share with tertiary education are higher than for any other clusters. For k-value=200, more than half of those aged 25-64 have tertiary education, this value drops below 0,50 only for k-values larger than 12,800. Yet, the affluence of U_ELI is perhaps most visible for income. For k-value=200 we can see that three out of ten of those aged 25-64 have an income in the top income decile. Even for k-value=51,200 the value of 0.20 puts U_ELI way ahead of all other clusters because for other affluent urban clusters the value for the share of top income earners does not exceed 0.15. The share of residents who receive social assistance is at the lowest level (below 0.03) up to k=12,800 when it is overtaken by cluster U_HOM. In terms of diversity, cluster five is characterized by moderate levels of European and non-European immigration at small scales, which increase gradually at larger scales.

Urban elite (U_ELI) residential areas are present in 42 out of 290 municipalities. Municipalities with the highest shares of such residential areas are: Täby, Lomma, Lidingö and Danderyd where over 90% of residents live in U_ELI residential areas. Three further municipalities have over half of residents living in U_ELI residential areas: Vellinge, Vaxholm and Nacka. These areas are part of North Eastern Stockholm and the coastal suburbs of Malmö. U_ELI is a highly isolated cluster, with isolation index equal to 0.85 (it has the second highest value of isolation index after U_HOM). When U_ELI residential areas share neighborhoods with other cluster types these tend to be U_HOM (0.04) or R_HOM (0.04).

Figure 42 Urban elite cluster, geographic distribution

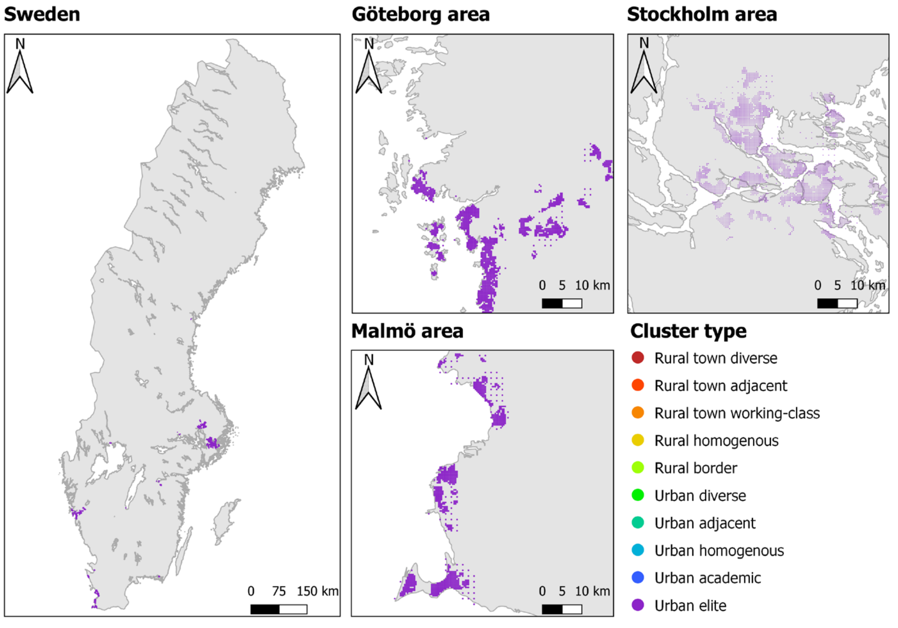
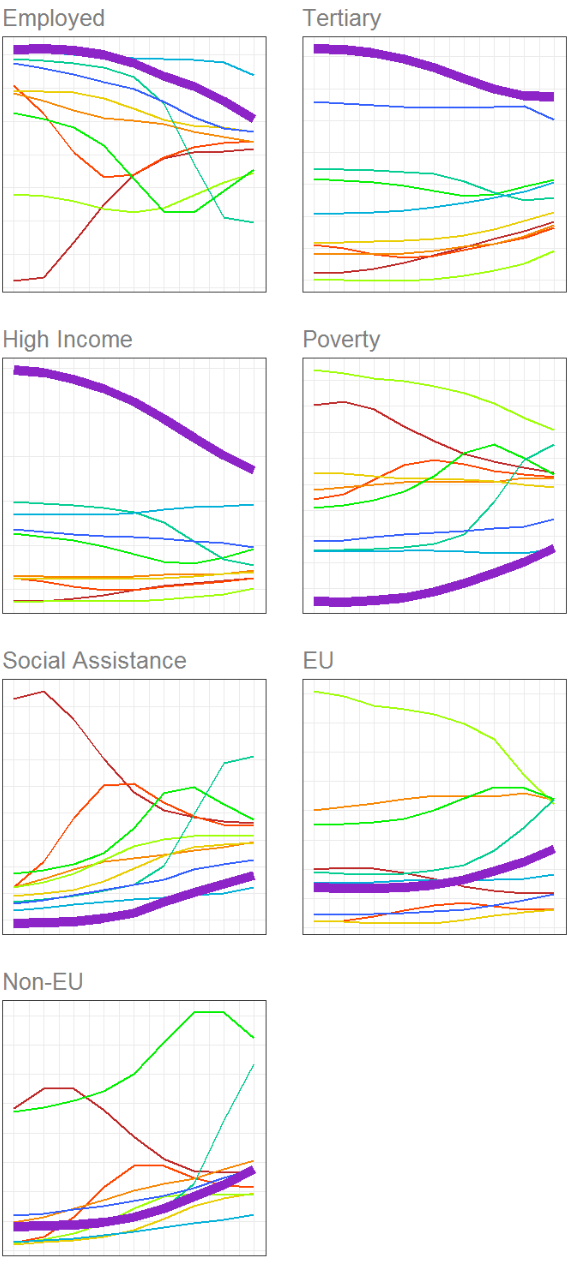


Figure 43 Urban elite cluster, by indicator



4. Geographic case studies

In this chapter we discuss three regions of interest and discuss clusters typical for residential areas in Gävle, Mälardalen and Värmland. Regions were chosen to cover all the cluster type and include middle size town, major metropolitan area, and less densely populated region with a border area.

4.1 The Gävle region

With exception of two cities (Gävle and Falun), Gävle region is a mostly rural region. Smaller settlements belong to *Rural homogenous* (R_HOM) cluster and small towns such as Hofors, Sandviken or Ockelbo feature residential areas belonging to *Rural town diverse* (RT_DIV) and *Rural town adjacent* (RT_ADJ) clusters. Hofors, Ockelbo or Skurskär present a typical pattern for small towns in rural municipalities where diverse areas are clustered around the center and are surrounded by *Rural town adjacent* (RT_ADJ) cluster. Interestingly, we see both *Rural town diverse* (RT_DIV) and *Urban diverse* (U_DIV) residential areas in Sandviken and Gävle, but not in Falun where there are only *Rural town diverse* (RT_DIV) cluster areas. This could relate to the scale of segregation and suggest that Gävle has larger scale segregation than Falun, but it is worth pointing out that they both belong to buffer subcluster or U_DIV and not the core. We also see a small number of *Rural town working-class* (RT_WC) areas which are concentrated in Långshytta locality in South Eastern Dalarna, an old industrial area which closed in recent years. In Gävle municipality (highlighted on the map) we see a combination of rural and urban cluster types. The largest cluster type is *Rural homogenous* (R_HOM) which accounts for 40% of residential areas, followed by *Rural town adjacent* (RT_ADJ) with 31% and *Urban adjacent* (U_ADJ) with 11%. The fourth largest type is *Rural town adjacent* (RT_ADJ) cluster with 10%. There are some *Urban academic* (U_ACA) residential areas in Falun and a smaller number in Gävle, these are to proximity to universities: Dalarna University and Gävle Högskola. In southern Gävle *Urban academic* (U_ACA) areas border many different cluster types: *Urban adjacent* (U_ADJ)*, Rural homogenous* (R_HOM)*, Rural town diverse* (RT_DIV)*, Urban diverse* (U_DIV) *and Rural town adjacent* (RT_ADJ).

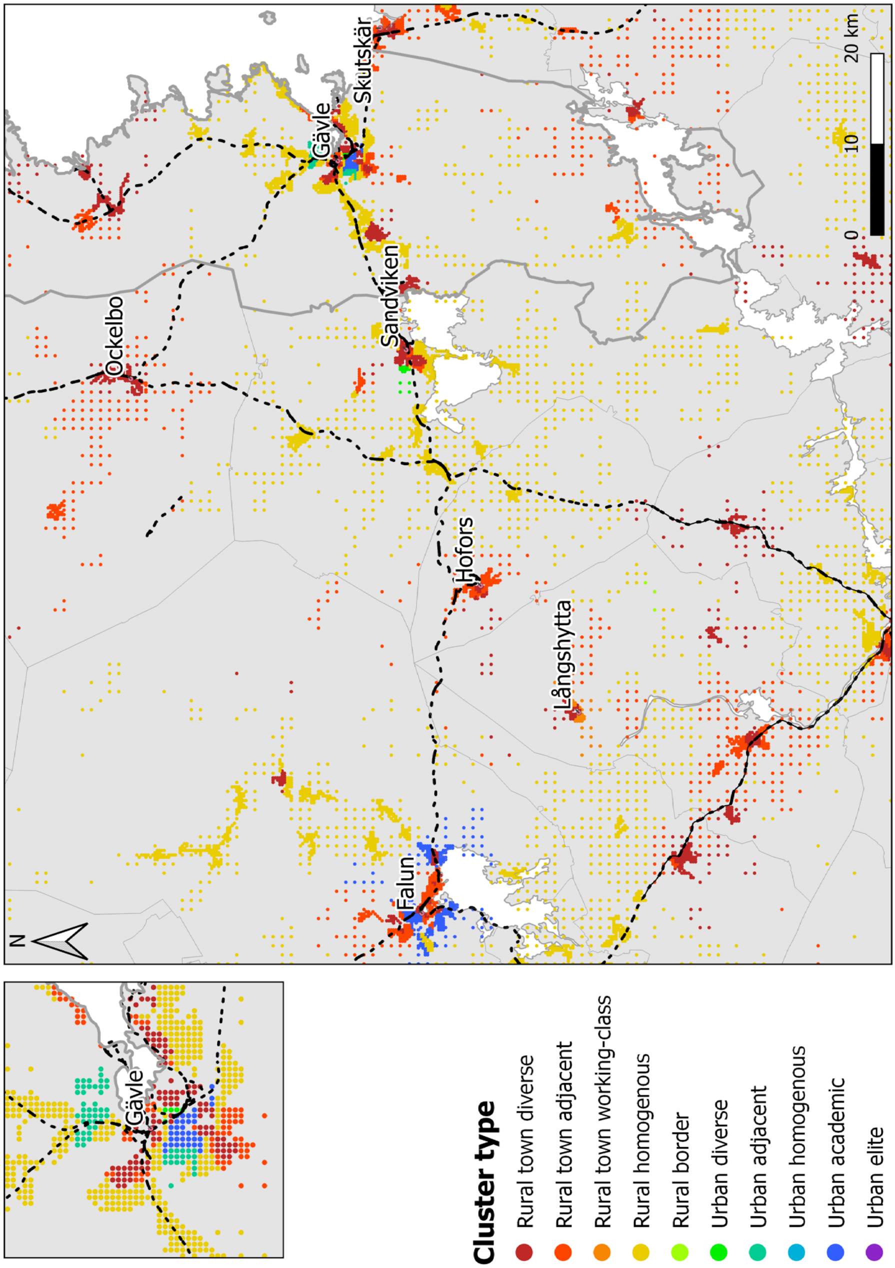


Figure 44 Geographic distribution of clusters in Gävle region

4.2 The Mälardalen region

Mälerdalen region contains an equal mix of rural and urban clusters. Stockholm municipality (highlighted on the map) has a large share of *Urban academic* (U_ACA) (37%), *Urban diverse* (U_DIV) (31%) and *Urban elite* (U_ELI) (28%) residential areas. Within Stockholm municipality these are some core subclusters of U_DIV, they are found in Rinkeby, Hässely Gård, Husby, Kista, Skärholmen or Hagsätra. They tend to border *Urban adjacent* (U_ADJ), *Urban academic* (U_ACA) but also *Urban elite* (U_ELI) cluster types. The high share of residential areas in *Urban diverse* (U_DIV) cluster is quite unique to the Stockholm region. This cluster can be found in Stockholm's suburban municipalities such as Södertälje or Botkyrka, and in Märsta. There are patches (U_DIV) in smaller towns like Eskilstuna, but these belong to buffer subcluster, not the core. In Stockholm we can distinguish large-scale *Urban elite* (U_ELI) areas such as Danderyd but also smaller scale Mälarhojden, an area located *between Urban academic* (U_ACA) cluster in Hägersten and *Urban diverse* (U_DIV) cluster in Bredäng. Large part of coastal areas north of Mälaren are classified as *Urban homogenous* (U_HOM) cluster, key features of this cluster are its high labor market attachment, low educational attainment, and low share of immigrants. This cluster tends to be buffered from diverse clusters with patches of *Urban adjacent* (U_ADJ) cluster. A typical example of *Urban adjacent* areas is Nykvarn, a rural locality with access to fast rail connection to Stockholm. Nykvarn is located 10 kilometers from Södertälje, a municipality which has one of the highest shares of foreign-born residents in Sweden. To the West of Nykvarn, there are residential areas belonging to *Rural town working-class* (RT_WK) cluster. RT_WK areas are common in Mälardalen region. We see several smaller localities with majority of RT_WK residential areas, especially on southern rail line (Järna, Mölnbo, Gnesta, Björnlunda) and in Åkers styckebruk, an old industrial town.

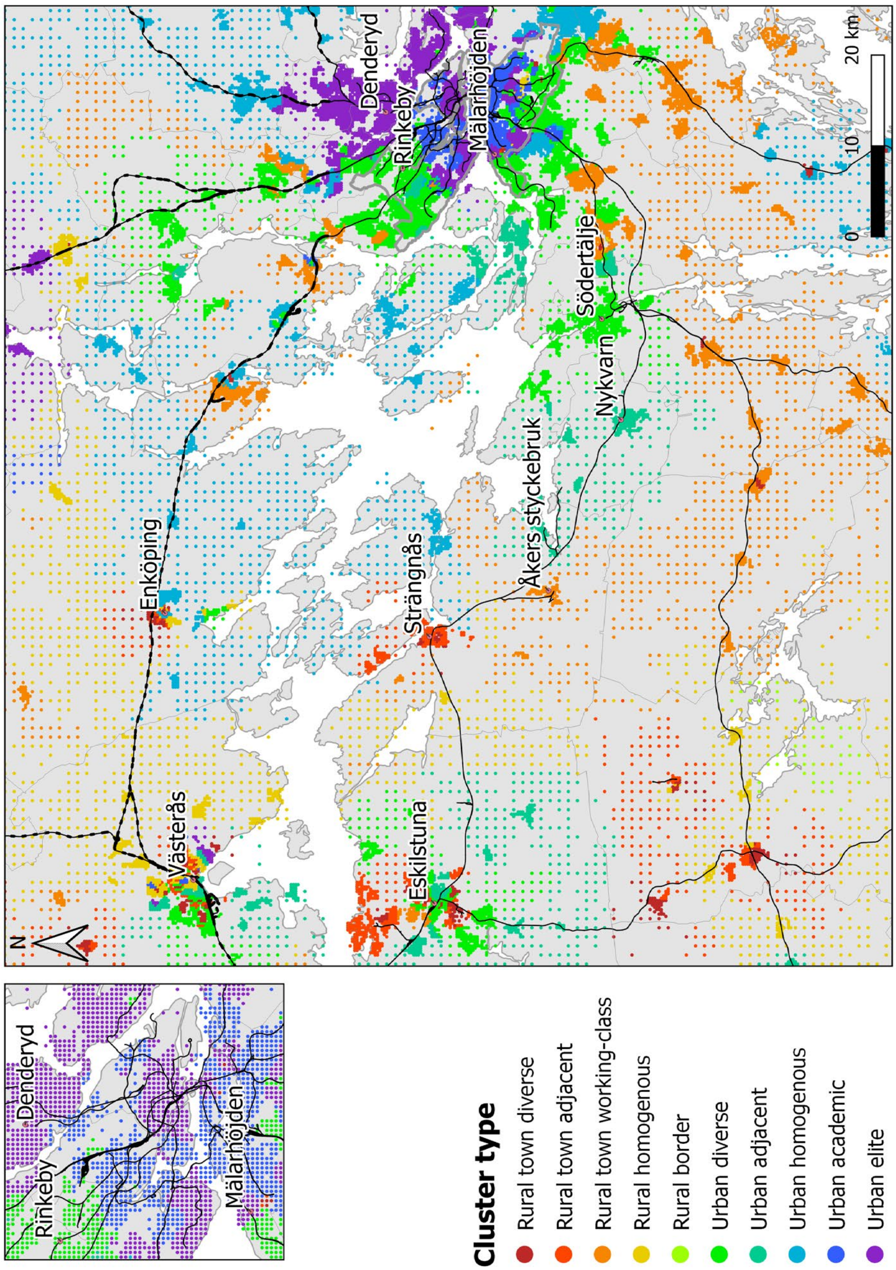


Figure 45 Geographic distribution of clusters in Mälardalen region

4.3 The Värmland region

Värmland region offers a blend of rural clusters, including a rare *Rural border area* (R_BOR) cluster. It also features a sizable concentration of *Urban academic* (U_ACA) cluster in Karlstad and Skoghall. Presence of these residential areas is likely related to three factors: proximity to Karlstad university, presence of industries which require skilled workers and advantageous geographical location, with proximity to lake Vänern. In Skoghall, a locality where chemical and paper industries were historically big employers, we see patches of *Rural town working-class* (RT_WC) residential areas. In contrast to previous two regions, in Värmland we can also find *Rural border area* (R_BOR) residential areas. In Årjäng, a locality 30 kilometers from the Norwegian border, R_BOR is the dominant cluster type. Several localities on the rail line (Arvika, Kil, Grums, Säffle, Åmål) follow a common pattern of small and mid-sized towns, with both *Rural town diverse* (RT_DIV) and *Rural town adjacent* (RT_ADJ) residential areas. Outside of towns the dominant types are *Rural border* (R_BOR) and *Rural homogenous* (R_HOM) cluster.

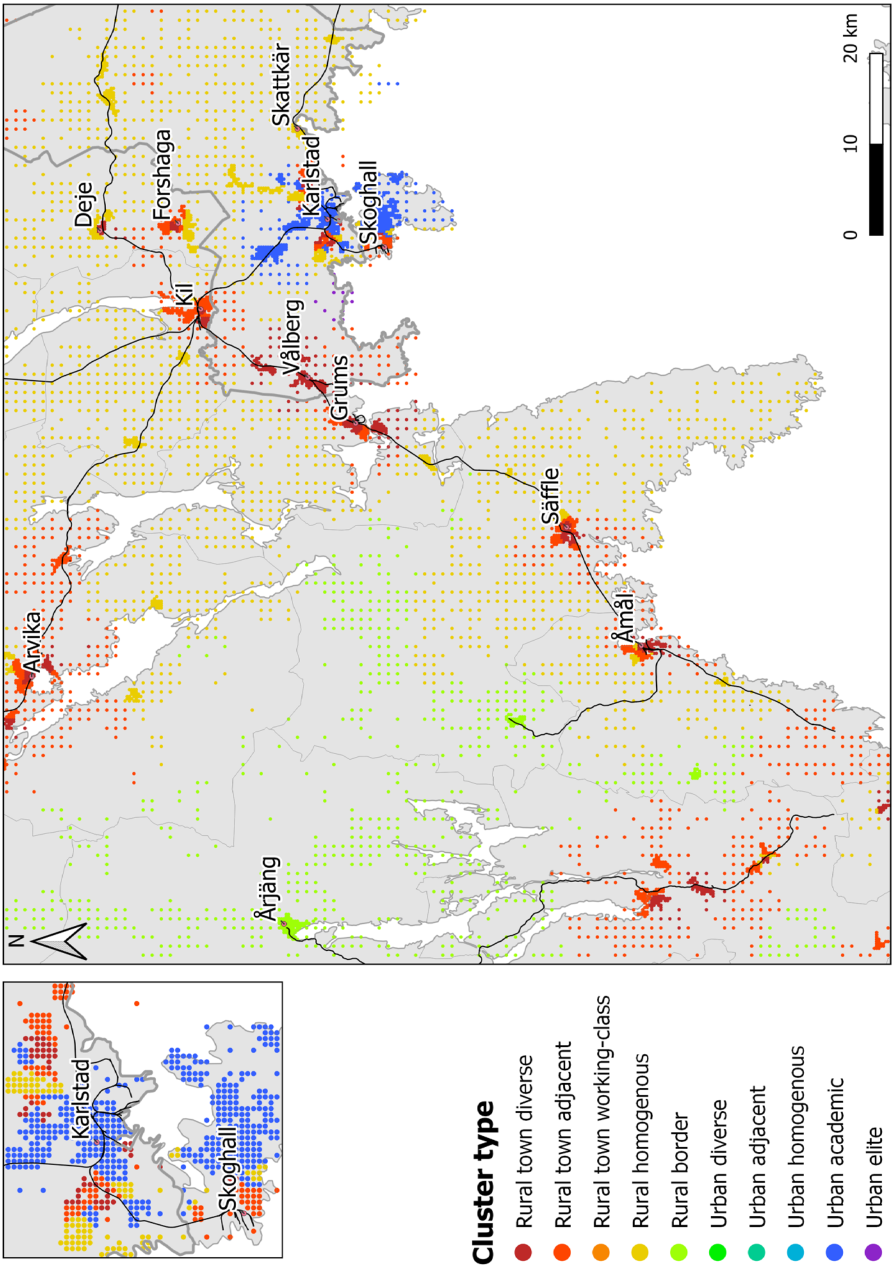


Figure 46 Geographic distribution of clusters in Värmland region

5. Multiscalar typology and other typologies

In this chapter we present a cross tabulation of multiscalar typology with other commonly used typologies. In 2018 Statistics Sweden (SCB) released a new regional division called DeSO (Demographic Statistics Areas), which replaced older SAMS. There are 5,984 such DeSO areas and their unique codes contain information about the municipality that they belong to as well as a letter which stands for area classification. Category A refers to areas outside of larger population concentrations and urban areas, category B is for a population concentration or urban areas outside of municipality central town, and category C is for municipality central towns. Table 6 shows a cross-tabulation of our classification and DeSO categories.

Table 6 Cross-tabulation of multiscalar topology and DeSO

Cluster	**A**	**B**	**C**
RT_DIV	0.3	0.22	**0.48**
RT_ADJ	**0.51**	0.06	0.43
RT_WC	**0.55**	0.13	0.32
R_HOM	**0.72**	0.12	0.16
R_BOR	**0.87**	0.04	0.09
U_DIV	0.15	0.07	**0.79**
U_ADJ	0.31	0.14	**0.56**
U_HOM	0.36	0.18	**0.46**
U_ACA	0.25	0.12	**0.63**
U_ELI	0.06	0.09	**0.84**

We see that with the exception of *Rural town diverse* (RT_DIV) our classification corresponds well to DeSO classification, that is all clusters that are classified as urban have majority of residential areas in DeSOs which are classified as C areas. Clusters which are rural or rural town have the majority of their residential areas in A category, that is areas outside of larger concentration and urban areas. *Rural town diverse* (RT_DIV) cluster is an exception because of the definition of category C, where it is defined not only by the population density criteria but also by whether the area is a part of municipality's seat or its central city. While category C contains DeSO areas in Stockholm, Goteborg, and Malmö, it also contains areas which may not be commonly considered as central or urban. For instance, the town of Årjäng, which is the seat of Årjäng municipality has less than 5,000 residents yet it is classified as DeSO area in C category. In our classification Årjäng is covered exclusively by *Rural border areas* (R_BOR) residential areas.

Another, more detailed, classification has been developed by Swedish Association of Local Authorities and Regions (SKR). The latest classification is from 2017 and it applies to municipal group divisions. This classification consists of nine groups divided into smaller subgroups. Municipalities are grouped according to their population, proximity to larger urban agglomerations and commuting patterns.

Major group A is called Large cities and municipalities near large cities. It consists of two subgroups: A1 Large cities and A2 Commuting municipalities near large cities. Major group B refers to Medium-sized town and municipalities near medium sized towns. It consists of three subgroups: B3 Medium-sized towns, B4 Commuting municipalities near medium sized towns and B5 Commuting municipalities with a low commuting rate near medium sized towns. Major group C is Smaller towns-urban areas and rural municipalities. Here, there are four subgroups: C6 Small towns, C7 Commuting municipalities near small towns, C8 Rural municipalities, and C9 Rural municipalities with a visitor industry. Table 7 shows cross-tabulation of our classification and SKR typology.

Table 7 Cross-tabulation of multiscalar topology and SKR typology

	Large cities and near large cities		Medium sized towns and near medium sized towns			Smaller towns, urban areas and rural			
Cluster	**A1**	**A2**	**B3**	**B4**	**B5**	**C6**	**C7**	**C8**	**C9**
RT_DIV	0.03	0.02	0.13	0.14	0.14	0.18	**0.19**	0.14	0.04
RT_ADJ	0	0.01	0.1	0.13	0.15	0.17	0.18	**0.22**	0.04
RT_WC	0.01	0.14	0.15	**0.2**	0.12	0.13	0.14	0.11	0
R_HOM	0	0.04	0.14	0.14	0.14	**0.25**	0.12	0.11	0.06
R_BOR	0	0	0	0.21	0.08	0.06	0.16	**0.34**	0.15
U_DIV	0.19	0.32	**0.41**	0.01	0.01	0.05	0	0	0
U_ADJ	0.23	0.13	**0.47**	0.17	0	0.01	0	0	0
U_HOM	0	**0.69**	0.01	0.04	0.07	0.1	0.06	0.03	0
U_ACA	0.12	0.06	**0.61**	0.06	0.02	0.1	0.04	0	0
U_ELI	0.25	**0.62**	0.09	0.03	0	0.01	0	0	0

Four cluster have the largest share of residential areas in group C: *Rural town diverse* (RT_DIV), *Rural town adjacent* (RT_ADJ), *Rural homogenous*

(R_HOM) and *Rural border* (R_BOR). The most rural cluster is *Rural border* (R_BOR) with 34% of residential areas belonging to C8 subgroup of Rural municipalities. *Rural town working-class* (RT_WC) cluster is classified as group B. 20% of residential areas in this cluster belong subgroup B4, Commuting municipalities near medium sized town. In group B there are also three urban clusters: *Urban diverse* (U_DIV), *Urban adjacent* (U_ADJ) and *Urban academic* (U_ACA). All fall in subgroup B3, medium-sized towns. *Urban academic cluster* (U_ACA) has especially high concentration in B3 group, 61%. Two clusters have majority of residential areas falling into group A of Large cities and municipalities near large cities: *Urban homogenous* (U_HOM) and *Urban elite* (U_ELI). One notable feature is that some clusters are evenly spread between subgroups (like *Rural town diverse, Rural town adjacent* or *Rural town working-class*) while most of urban clusters are concentrated in a small number of subgroups. With SKR classification we also see the difference between *Rural town diverse* (RT_DIV) and *Urban diverse* (U_DIV), two diverse clusters. The first one has at least 10% of residential areas in six subclusters in groups B and C. The second is concentrated in only three subclusters (two falling in group A) and has hardly any residential areas in group C of Smaller towns/urban areas and rural municipalities.

6. Conclusions and Discussion

In this study we have classified populated grid cells (250-m squares in urban areas and 1,000-m squares in rural areas) in Sweden on the basis of the socio-demographic composition of their neighborhood population in 2016 using indicators of high income, education, employment, social allowances, at-risk-of-poverty, and migration status. We have used multi-scalar measures of geographical context which implies that the classification is influenced both by the population composition of the immediate neighborhood and by the population composition of the larger local area.

We have found that there exists one set of neighborhood types that are found in rural, or non-metropolitan areas, and another set of clusters that are found in urban, primarily metropolitan areas. Our classification maps well into existing urban/rural classifications but offers a finer resolution and a more detailed description based on multiple domains[3].

With respect to the urban clusters, what we have found corresponds in broad terms to what one would expect. Primarily, the urban cluster demonstrate clear patterns of spatial polarization. At one end there is the *Urban elite* cluster (U_ELI), characterized by high proportions of high-income individuals and high proportions of individuals with a tertiary education. At the other end, one finds the *Urban diverse* (U_DIV) cluster, with few high-income individuals, and high proportions of households at risk of poverty receiving social allowances. This cluster also is characterized by high shares of non-European immigrants and high share of European immigrants. Between these poles one finds the *Urban academic* (U_ACA) cluster which is similar to the *Urban elite* (U_ELI) but with smaller proportions of high-income individuals. Separated from the *Urban diverse* (U_DIV) cluster one also finds the *Urban adjacent* (U_ADJ) cluster with high levels of employment, tertiary education, and high income at lower spatial scales, but lower employment, increasing rates of poverty, social allowances, and non-European immigrants at larger neighborhood scales. Even more removed from the central parts of the metropolitan areas, one finds the *Urban homogenous* (U_HOM) cluster with fewer immigrants, low levels of poverty, and medium levels of tertiary education. In some ways, this cluster is similar to the rural clusters, which fits well with its location farther away from the metropolitan core. This composition of clusters implies

[3] For example, we can show that even though socio-economic deprivation tends to be associated with diversity (and especially non-European immigration), there are some clusters which have few high-income earners or low educational attainment despite relative absence of immigration.

that metropolitan areas are characterized by a diversity of neighborhood types across relatively short distances. That is, there are sharp contrasts to be found if one would navigate across the metropolitan landscape. "The city consequently tends to resemble a mosaic of social worlds in which the transition from one to the other is abrupt" (Wirth 1938).

The rural clusters we have found, on the other hand, did not fully correspond to our expectations. Judging from, for example, (Andersson et al. 2019) we would have expected clusters that represented neither elite concentrations, nor concentrations of vulnerable groups. Instead, one important cluster we found was *Rural town diverse*, with a profile that in many ways corresponds to neighborhoods of concentrated poverty found in metropolitan context: low levels of employment, high levels of social allowances and of poverty, and high proportions of non-European immigrants. Given clusters based on multiscalar measures these poverty-concentrations also put their mark on *Rural town adjacent*, and to some extent on *Rural town homogenous*. Our interpretation of this finding is that it reflects the large inflow of non-European immigrants that Sweden has experienced in the 2000s, and inflow that to an increasing extent has been directed to non-metropolitan areas. This interpretation is also supported by our analysis of the age of arrival of migrants. Another surprising cluster is *Rural border* which has high levels of poverty not matched by similarly high levels of social assistance. It is possible that some residents of these border zones, many of whom are born on the other side of the border, earn their income outside of Sweden, hence distorting poverty indicators for these areas.

Regarding limitations, our classification should be understood similarly to Weberian ideal types, i.e. clusters represent average profiles grid cells which are most representative for a given cluster type. Even though we can assign each grid cell to one of the ten clusters, individual grid cells and their profiles will deviate from their ideal type. We can estimate how well each observation fits in its assigned cluster type by looking at its silhouette value. This way we can separate the "most representative" observations for each type.

Neighborhood contexts shape life outcomes of those embedded in them. Research on neighborhood effects has often used predefined administrative units which treat neighborhoods as independent of surrounding areas. Yet, the way people experience their residential context depends not only on characteristics of their most proximate neighbors but relates to several features which can only be captured with multiscale approaches: proximity of ethnic cliffs, living in ethnic enclaves, concentrated poverty or affluence which persists over large scales. Our classification not only gives an updated overview of the types of neighborhoods that can be observed in Sweden in 2016 but also offers ample opportunities for future research to uncover how different neighborhood types relate to attitudes, preferences, and experiences of its residents.

References

Aldskogius, H. (1982) Swedish Urban Dimensions in the 1970's. *Geografiska Annaler: Series B, Human Geography*, 64, 77-87.

Andersson, E. K., M. Abramsson & B. Malmberg (2019) Patterns of changing residential preferences during late adulthood. *Ageing & Society*, 39, 1752-1781.

Andersson, E. K., T. H. Lyngstad & B. Sleutjes (2018) Comparing Patterns of Segregation in North-Western Europe: A Multiscalar Approach. *European Journal of Population.*

Andersson, E. K. & B. Malmberg (2015) Contextual effects on educational attainment in individualised, scalable neighbourhoods: Differences across gender and social class. *Urban Studies*, 52, 2117-2133.

--- (2018) Segregation and the effects of adolescent residential context on poverty risks and early income career: A study of the Swedish 1980 cohort. *Urban Studies,* 55, 365-383.

Andersson, E. K., T. Wimark & B. Malmberg (2020) Tenure type mixing and segregation. *Housing Studies*, 1-24.

Bell, W. (1953) The social areas of the San Francisco Bay region. *American Sociological Review*, 18, 39-47.

--- (1955) Economic, family, and ethnic status: An empirical test. *American Sociological Review*, 20, 45-52.

Bischoff, K. & S. F. Reardon. 2014. Residential segregation by income, 1970–2009. In *Diversity and disparities: America enters a new century*, ed. J. R. Logan, 208–234. New York, NY: Russell Sage.

Biterman, D. & E. Franzén (2007) Residential segregation (chapter 6). *International Journal of Social Welfare*, 16, 127-162.

Burgess, E. W. (1927) *The Urban Community: Selected Papers from the Proceedings of the American Sociological Society*, 1925. Chicago, University of Chicago.

Butler, T. & C. Hamnett (2012) Social geographic interpretations of housing spaces. Sage *handbook of housing studies*, 147-162.

Chang, F., W. Qiu, R. H. Zamar, R. Lazarus, & X. Wang,(2010). clues: An R Package for Nonparametric Clustering Based on Local Shrinking. *Journal of Statistical Software*, 33(1), 1–16.

Chetty, R. & N Clark, W. A. V. (1991) Residential preferences and neighborhood racial segregation - a test of the Schelling segregation model. *Demography*, 28, 1-19.

Hendren (2018) The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects. *The Quarterly Journal of Economics*, 133, 1107-1162.

Chetty, R., N. Hendren, P. Kline & E. Saez (2014) Where is the land of opportunity? The geography of intergenerational mobility in the United States. *The Quarterly Journal of Economics*, 129, 1553-1623.

Clark, W. A. V. (1991) Residential preferences and neighborhood racial segregation - a test of the schelling segregation model. *Demography*, 28, 1-19.

Costa, R., B. Malmberg, K. Haandrikman, A. F. Rogne & B. Sleutjes. (2018). Socio-economic Segregation in European Cities A Comparative Study of Amsterdam, Oslo, Stockholm and Copenhagen. In *ENHR 2018,* Uppsala, Sweden.

Dahlbäck, O. (2016). *Environment and crime among residents in urban areas: A study of districts in Stockholm*. Routledge.

Graham, B. S. (2018) Identifying and estimating neighborhood effects. *Journal of Economic Literature*, 56, 450-500.

Hennerdal, P. (2019). Geocontext [Python]. https://github.com/PonHen/geocontext

Janson, C.-G. (1971) A preliminary report on Swedish urban spatial structure. *Economic geography*, 47, 249-257.

--- (1980) Factorial social ecology: An attempt at summary and evaluation. *Annual Review of Sociology,* 6, 433-456.

Kaufman, L., & P. J. Rousseeuw (2005). *Finding Groups in Data: An Introduction to Cluster Analysis* (1st edition). Wiley-Interscience.

Malmberg, B. & E. K. Andersson (2019) Adolescent neighbourhood context and transition to parenthood: A longitudinal study. *Population, space and place*, 25, e2228.

Malmberg, B., E. K. Andersson & Z. Bergsten (2014) Composite Geographical Context and School Choice Attitudes in Sweden: A Study Based on Individually Defined, Scalable Neighborhoods. *Annals of the Association of American Geographers*, 104, 869-888.

Malmberg, B., E. K. Andersson & T. Wimark. 2018. The housing tenure landscape that structures Swedish society: Patterns and change 1990 to 2012 In *AAG 2018*. New Orleans, USA.

Malmberg, B. & W. A. Clark (2020) Migration and Neighborhood Change in Sweden: The Interaction of Ethnic Choice and Income Constraints. *Geographical Analysis.*

Malmberg, B., M. M. Nielsen, E. Andersson & K. Haandrikman. 2016. Migration, Residential Mobility and Housing Policy: Residential segregation of European and non-European migrants in Sweden 1990-2012 In *ENHR 2016, Belfast,* UK.

Nielsen, M. M. & P. Hennerdal (2019) Segregation of Residents with Tertiary Education in Sweden from 1990 to 2012. *The Professional Geographer*, 71, 301-314.

Openshaw, S. (1984). The modifiable areal unit problem, CATMOG 38. *Geo Abstracts*, Norwich, UK.

Park, R. E. 1926. The urban community as a spatial pattern and a moral order. In *The urban community : selected papers from the Proceedings of the American Sociological Society*, 1925, ed. E. W. Burgess, 3-18. Chicago: University of Chicago Press.

Pred, A. R. (1984) Place as historically contingent process: Structuration and the time-geography of becoming places. *Annals of the Association of American Geographers*, 74, 279-297.

Sampson, R. J (2012). *Great American City: Chicago and the Enduring Neighborhood Effect*, Chicago, University of Chicago Press.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of mathematical sociology*, 1(2), 143-186.

Shevky, E. & W. Bell (1955) *Social area analysis; theory, illustrative application and computational procedures*. Stanford, CA: Stanford University Press

Shevky, E. & M. Williams. 1949. *The Social Areas of Los Angeles, Analysis and Typology*. University of California Press, Berkeley and Los Angeles.

Simonsen, K. 2016. Place as encounters: Practice, conjunction and co-existence. In *Mobility and Place,* 29-42. Routledge.

Sundquist, K., M. Malmstrom & S. E. Johansson (2004) Neighbourhood deprivation and incidence of coronary heart disease: a multilevel study of 2.6 million women and men in Sweden. *Journal of Epidemiology and Community Health*, 58, 71-77.

Toft, M. 2018. *Biographies of privilege: Spatiotemporal structures of upper-class formation*. Department of Sociology and Human Geography Faculty of Social Sciences, Oslo University, Oslo.

Vetenskapsrådet. 2018. Svensk forskning om segregation – en kartläggning.

Vogiazides, L. (2018) Exiting distressed neighbourhoods: The timing of spatial assimilation among international migrants in Sweden. *Population, Space and Place*, 24, e2169.

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58, 301, 236–244.

Warming, E., I. Bayley Balfour, P., Groom & M Vahl (1909) *Oecology of Plants: An Introduction to the Study of Plant-communities*. Oxford: Clarendon

Westergaard, J. H. (1970) Research Committee on Urban Sociology of the International Sociological Association: Report on a meeting in Stockholm in September 1969. *Social Science Information*, 9, 35-45.

Wikstrom, P.-O. H. (1991) *Urban crime, criminals, and victims: The Swedish experience in an Anglo-American comparative perspective.* NY: Springer-Verlag.

Wimark, T., K. Haandrikman & M. M. Nielsen (2019) Migrant labour market integration: The association between initial settlement and subsequent employment and income among migrants. *Geografiska Annaler: Series B, Human Geography*, 101, 118-137.

Wirth, L. (1938) Urbanism as a Way of Life. *The American Journal of Sociology*, 44, 1, 1-24.

Östh, J., B. Malmberg, & E. K. Andersson (2014). Analysing segregation using individualised neighbourhoods. In *Social-spatial segregation: Concepts, processes, and outcomes* (135–162). Policy Press.

Appendix 1 Rural clusters across Sweden

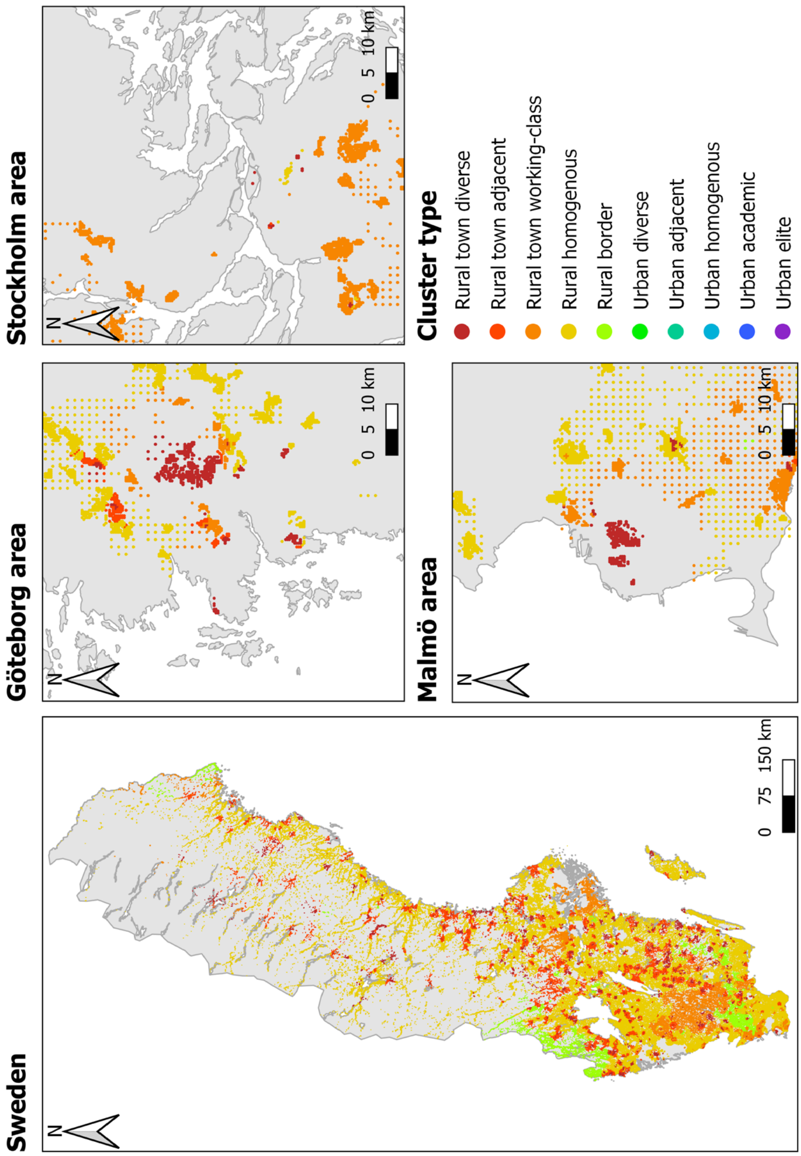


Figure 47 Rural clusters, geographic distribution

Appendix 2 Urban clusters across Sweden

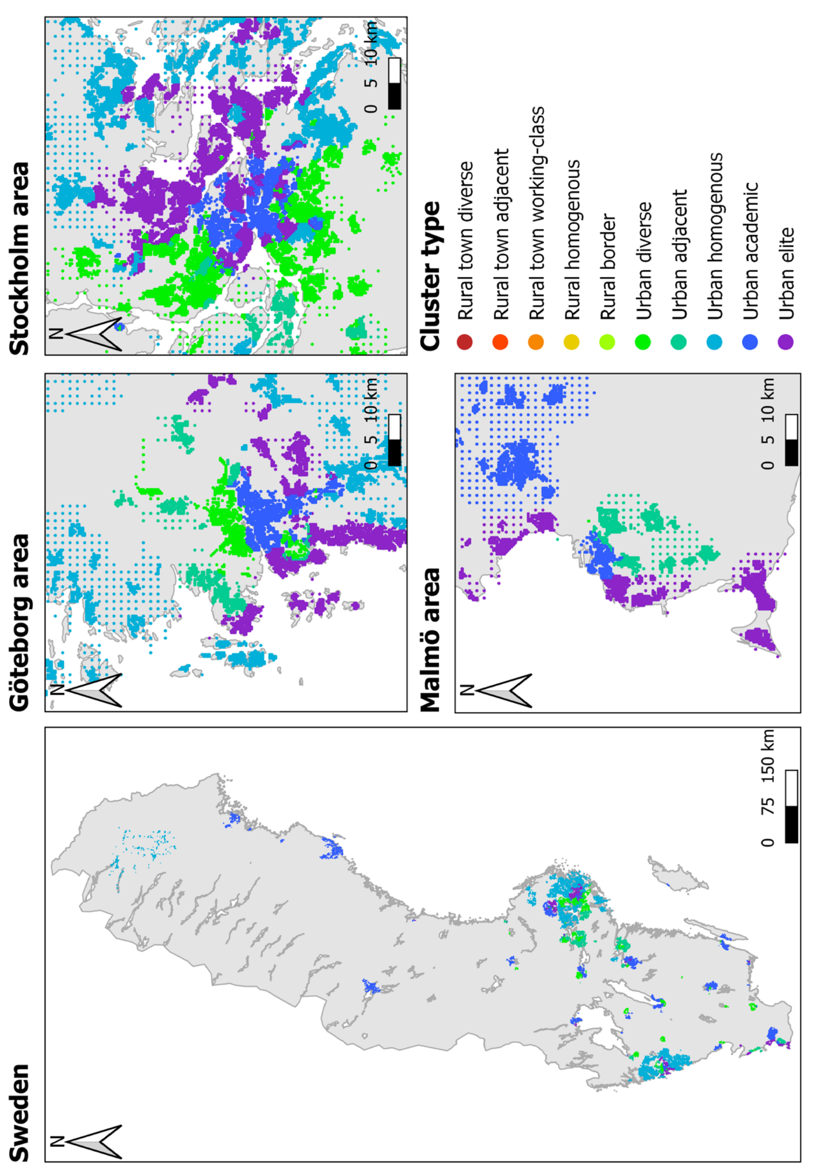


Figure 48 Urban clusters, geographic distribution

Appendix 3 Age by cluster

Table 8 Age, by cluster

Cluster	**Mean**	**S.D**
RT_DIV	43.77	24.41
RT_ADJ	41.29	24.10
RT_WC	39.63	24.62
R_HOM	42.68	24.31
R_BOR	44.52	24.46
U_DIV	38.01	22.99
U_ADJ	40.80	24.22
U_HOM	40.76	23.93
U_ACA	39.14	22.61
U_ELI	39.57	23.71

Appendix 4 Residential areas by cluster

Table 9 Share of residential areas and population, by cluster

Cluster	Residential Areas	Population
RT_DIV	0.08	0.14
RT_ADJ	0.13	0.09
RT_WC	0.10	0.09
R_HOM	0.44	0.21
R_BOR	0.04	0.01
U_DIV	0.03	0.11
U_ADJ	0.02	0.04
U_HOM	0.07	0.06
U_ACA	0.05	0.17
U_ELI	0.03	0.08

Appendix 6 Isolation and exposure indices

Table 10 Isolation and exposure indices

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) RT_DIV	**0.54**	0.18	0.05	0.19	0.01	0.01	0	0	0.02	0
(2) RT_ADJ	0.12	**0.56**	0.02	0.26	0.02	0.01	0	0	0.01	0
(3) RT_WC	0.04	0.03	**0.75**	0.11	0.02	0.02	0.01	0.01	0	0
(4) R_HOM	0.04	0.07	0.03	**0.82**	0.01	0	0	0.01	0.02	0
(5) R_BOR	0.03	0.06	0.07	0.15	**0.7**	0	0	0	0	0
(6) U_DIV	0.02	0.02	0.06	0.05	0	**0.76**	0.04	0.02	0.02	0.01
(7) U_ADJ	0.02	0.01	0.04	0.07	0	0.06	**0.77**	0.01	0	0.01
(8) U_HOM	0	0.01	0.02	0.05	0	0.01	0	**0.89**	0	0.02
(9) U_ACA	0.02	0.02	0	0.12	0	0.01	0	0	**0.8**	0.02
(10) U_ELI	0.01	0	0	0.04	0	0.01	0.01	0.04	0.03	**0.85**

Appendix 5 Subclusters of U_DIV cluster

When we inspected silhouette statistics for 10-cluster solution, we realized that *Urban diverse* (U_DIV) cluster had the lowest quality of clustering. We explored a possibility to distinguish between core and buffer zones for this cluster.

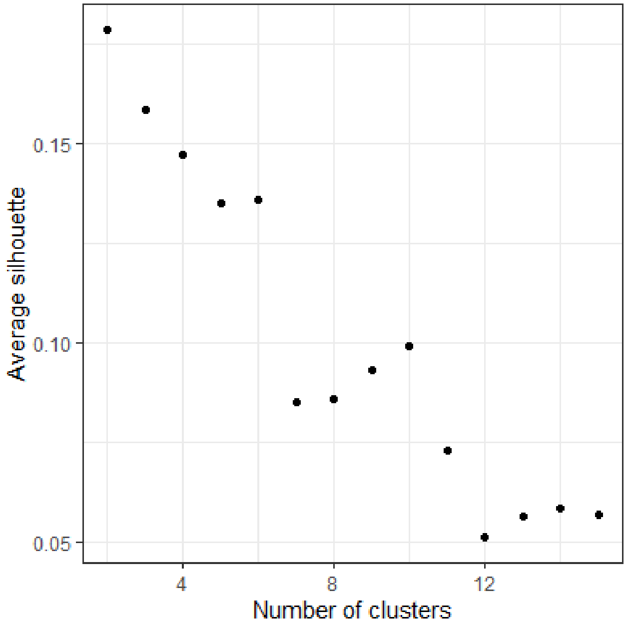


Figure 49 Average silhouette width for different cluster solutions

Choosing the optimal number of clusters is a fundamental issue in cluster analysis. One method is to pick a cluster solution with the highest average silhouette values. In our case this would lead to a solution with only two clusters, which does not capture the variability in residential area types. A slightly modified method is similar to the so-called elbow method. Here, we search for a point with sufficiently many clusters where additional more clusters to such solution leads to a large drop in the average silhouette value.

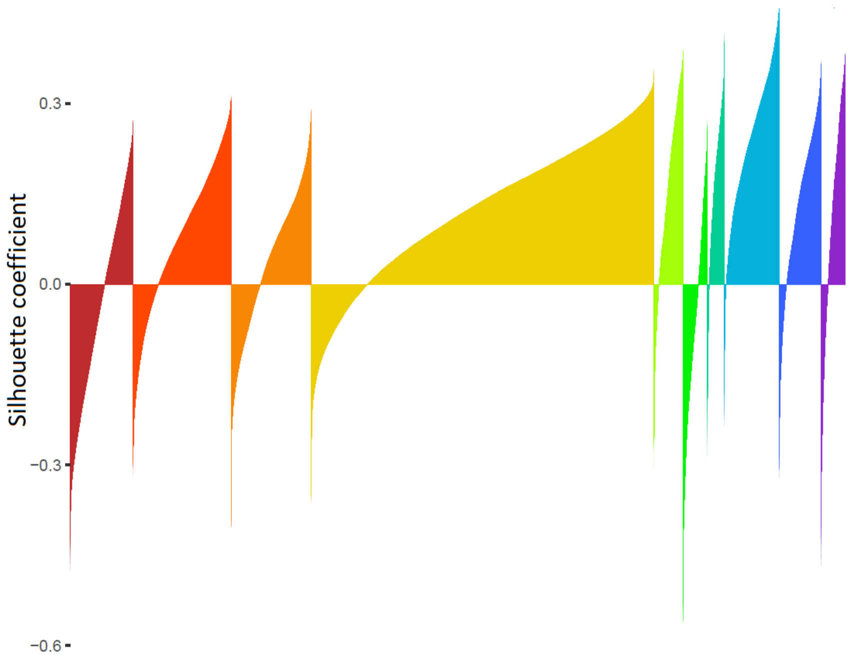


Figure 50 Silhouette coefficient for residential areas, by cluster

Figure 50 shows silhouette plots with individual silhouette coefficients for each observation in a ten-cluster solution. Silhouette coefficients measure how similar an observation is to other observations in its own cluster as opposed to observations in the neighbouring cluster. The values go between 1 and -1, with higher values indicating that an observation is well clustered and low values that it is poorly clustered. As can be seen, there are two problematic clusters in 10-cluster solution: "*Rural town diverse*" (RT_DIV) *and "Urban diverse"* (U_DIV). In both clusters there are many observations with negative silhouette coefficients. A negative value for silhouette coefficient means that this observation is not well clustered and that we should treat such observations with caution and not as representative cases for our proposed ideal types. Figure 51 shows the average silhouette value for each cluster in 10-cluster solution. Again, we see that cluster 1 (RT_DIV) and cluster 6 (U_DIV) have negative average silhouette width values. To address the problem of U_DIV we examined our hierarchical clustering and looked for a cluster solution where U_DIV breaks down into two separate clusters. This happens in a cluster solution with 14 clusters, where U_DIV in 10-cluster solution breaks into U_DIV buffer and U_DIV core in 14-cluster solution. Further analysis shows that U_DIV (noted as cluster 6 in Figure 51) breaks into U_DIV buffer which is a buffer zone for urban deprivation (*outer subcluster*) and U_DIV core with high average silhouette value which represents core urban deprivation (*inner subcluster*). Subclusters of U_DIV are marked as cluster 9 and cluster 10 in Figure 52.

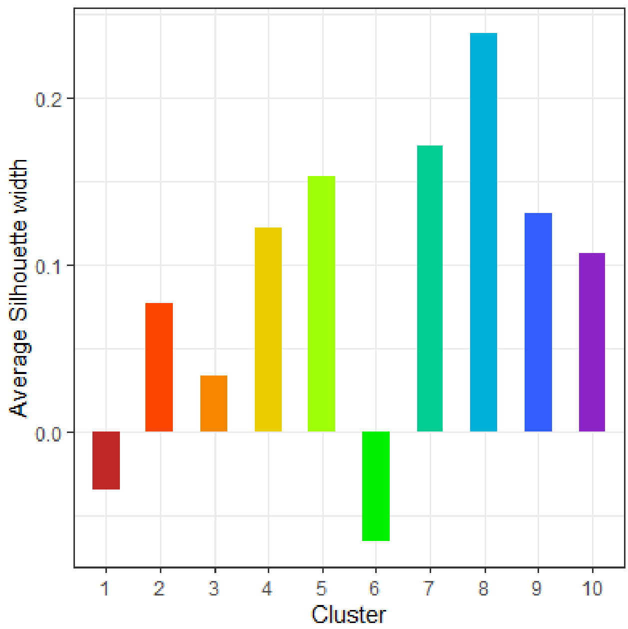


Figure 51 Average silhouette width, by cluster for 10-cluster solution

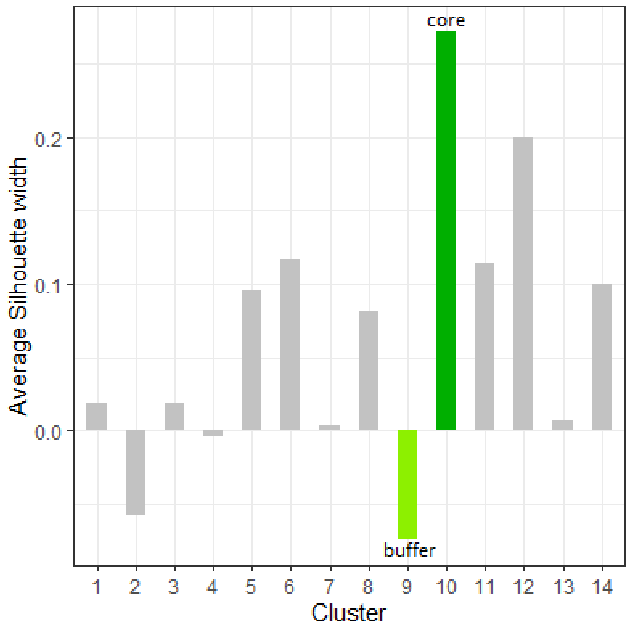


Figure 52 Average silhouette width, by cluster for 14-cluster solution

Figure 53 Urban diverse subclusters, geographic distribution

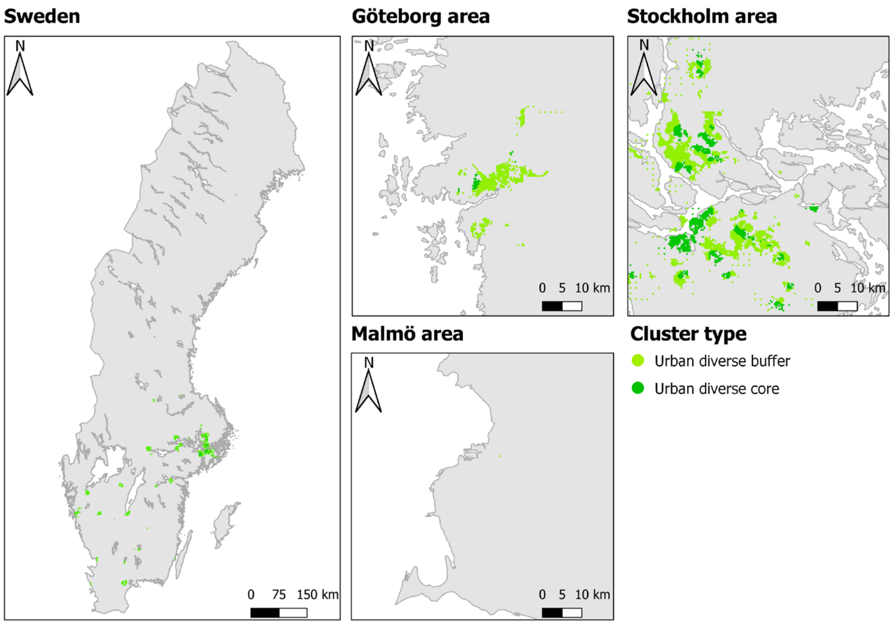


Figure 54 Urban diverse subclusters, by indicator