

# Spatial segregation: a data science approach

Nick van Gils  
September 2018

## Summary

### *Introduction*

From 2001 to 2015 income inequality in the Netherlands remained largely unchanged. The low level of income inequality in the recent history of the Netherlands, is often seen as a result of the functioning of the welfare state. However, recent developments such as the aging population and globalization might lead to a decreasing role of the welfare state, resulting in an increasing gap between economically strong and economically weak layers of the population. One of the factors affecting the gap between population groups is spatial segregation. If people of different social groups live and spend their time in different geographical areas, they have less chance to become acquainted and develop an understanding for each other.

High levels of segregation are often thought to have negative effects on minorities and lower income groups. However, people in a residentially segregated city might still mingle during daily activities, such as visiting restaurants, visiting a pub, attending school and shopping, this is referred to as activity segregation.

The research question for this thesis is formulated as, *“How large is spatial segregation in the Netherlands, both in residential location and in the location of activities?”* The aim of this thesis is to provide further insights into spatial segregation, specifically for five most populous cities in the Netherlands. This is done by analyzing the two forms of segregation: residential segregation and activity segregation. For the analysis of activity segregation Amsterdam is taken as a case study. For this thesis the term segregation is defined as an overrepresentation of specific population groups in streets, neighborhoods or districts in comparison to the average of a city or state.

### *Literature review*

Different studies show that different characteristics can be used to group people and a number of these groupings are used in this study to analyze residential segregation. These groupings are presented along the lines of four themes in this thesis, (I) ethnicity, (II) opportunities, (III) housing, (IV) natural processes.

Residential Segregation can be caused by a natural process, this process describes population changes within an area, which is not happening in the rest of the city. Selective migration plays an important role in the level of segregation and change in segregation level. In the literature selective migration is seen as the result of preferences of households and (the lack of) possible choices.

### *Methodology*

Residential segregation is seen as a multidimensional phenomenon, therefore a variety of indices exist to measure it. However, only a limited number of these are often applied in analyses. A well established and well known measure, the dissimilarity index, is used in this thesis as to measure segregation, supported by a visual representation (dot maps) of the different population groups. Together these two tools can give sufficient insight into the distribution of people in the five cities to ultimately determine if significant segregation can be found.

To measure the levels of segregation census data is used. The cities that are studied are the five most populous cities in the Netherlands, and together they house around 15% of the country's population. The demographic composition of the five cities is different from that of the country. In the cities multiple population characteristics are overrepresented such as immigrants, low income and one person households.

The second part of this thesis focusses on the activity segregation in Amsterdam. Using a relatively new method of segregations study, a first attempt is made to quantify the segregation of activity in Amsterdam and to figure out how much of this segregation can be explained by (I) residential segregation and (II) people experiencing a social barrier to visiting neighborhoods with another demographic (ethnic) composition. To explain motive behind activity segregation, a so called alternative-specific conditional logit model was used. For this model the OVIN dataset is used providing data on peoples movements in the Netherlands. Two motives are used from this dataset, social recreational visits, and visiting friends or family.

### *Results*

The results of analyzing the residential segregation are presented along the four themes mentioned above. In the first theme ethnicity, the non-western population group seemed the most segregated producing a moderate score according to the dissimilarity index in all five cities. When splitting the non-western population in smaller ethnic groups the Turkish and Moroccan population seemed most unevenly distributed in all five cities. For the opportunities theme only the high income population of The Hague produced a moderate segregation level, and low index scores were found for all other income groups. Indicating that income groups are evenly distributed in all five cities. The third theme housing produced the highest levels of segregation indicating that the housing market contributes to segregation in the five cities. Of all three tenure types, people living in social housing are most segregated. And the ratio between single- and multifamily homes produce the high levels of segregation. The fourth theme focusses on characteristics related to natural processes (age and household composition). A low index score was found for these characteristics indicating little to no segregation. The segregation levels were calculated for the period 2004 until 2017. The results did not suggest any large shifts in segregation level. The distribution of people and thus the segregation levels seem to be quite stable over the observed period.

After analyzing the residential segregation in the five cities, the activity segregation is studied in Amsterdam. A visual analysis of the ethnic groups' residential location and activity locations showed similar patterns indicating that people stay close to their home location for the activities in the dataset. The estimations produced by the logit models showed that both the spatial and the social barriers are important determinants of the activity location choice. So it seems that in Amsterdam people go out near to where they live, and they prefer locations with a similar demographic composition. To test the robustness of these results different definitions of the social barrier are used. Also studied is how much the destination characteristics affect the activity location choice. This analysis showed that users prefer to visit tracts with a low share of low income population and a low share of non-western inhabitants. A possible reason is that these tracts have fewer attractive amenities and facilities. Non-western users however seem to have little aversion toward these areas. Two amenities of the census tract were tested in the estimation, namely the amount of restaurants and the share of open and green space per census tract. The estimation showed that the amount of restaurants has a strong positive effect on people's preference to choose the area. The share of green space however produced a negative effect indicating that the social recreational activities in the estimation do not depend on green and open spaces.