



**HAL**  
open science

## Multidimensional urban segregation: an exploratory case study

Marie Cottrell, Aurélien Hazan, Madalina Olteanu, Julien Randon-Furling

### ► To cite this version:

Marie Cottrell, Aurélien Hazan, Madalina Olteanu, Julien Randon-Furling. Multidimensional urban segregation: an exploratory case study. 12th International Workshop on Self-Organizing Maps and Learning Vector Quantization, Clustering and Data Visualization (WSOM+) th International Workshop on Self-Organizing Maps and Learning Vector Quantization, Clustering and Data Visualization (WSOM+), Jun 2017, Nancy, France. hal-01519723

**HAL Id: hal-01519723**

**<https://hal.science/hal-01519723v1>**

Submitted on 28 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Multidimensional urban segregation: an exploratory case study

Marie Cottrell  
Madalina Olteanu  
Julien Randon-Furling  
Université Paris 1 Panthéon-Sorbonne  
SAMM (EA4543)  
Paris, France

Email: marie.cottrell, madalina.olteanu, julien.randon-furling@univ-paris1.fr

Aurélien Hazan  
UPEC - Université Paris-Est Créteil  
LISSI  
Créteil, France  
Email: aurelien.hazan@u-pec.fr

**Abstract**—Segregation phenomena have long been a concern for policy makers and urban planners, and much attention has been devoted to their study, especially in the fields of quantitative sociology and geography. Perhaps the most common example of urban segregation corresponds to different groups living in different neighbourhoods across a city, with very few neighbourhoods where all groups are represented in roughly the same proportions as in the whole city itself. The social groups in question are usually defined according to one variable: ethnic group, income category, religious group, electoral group, age... In this paper, we introduce a novel, multidimensional approach based on the Self-Organizing Map algorithm (SOM). Working with public data available for the city of Paris, we illustrate how this method allows one to describe the complex interplay between social groups’ residential patterns and the geography of metropolitan facilities and services.

## I. INTRODUCTION

Various contexts can be characterized as reflecting one form or another of urban segregation. In most instances, the phenomenon is defined to be one where the distribution of a given variable locally (at the neighbourhood scale) differs substantially from its distribution at the scale of the whole city. A seminal study of the emergence of segregation was Schelling’s checker-board simulation of a two-community dynamics [1], [2] – incidentally, one of the first multi-agent models. This model revealed how, even with relatively high levels of tolerance for mixity, a system of agents located at sites on a grid would eventually evolve towards patterns of segregation, with the two communities “living” in distinct parts of the grid. Schelling’s model has given rise to a vast literature on segregation across a wide range of disciplines [3]–[14], and notably in statistical physics, where analogies with certain interacting particle systems have been put to fruitful use [15]–[24]. However, rare are the instances where real-world data has been compared with theoretical results other than stylized facts [25]–[27]. One reason for that is the notorious difficulty to elaborate, based on available data, segregation indices [28] that could correspond to some of the theoretical ones – the evolution of which is described in agent-based models.

We adopt a somewhat converse strategy: starting from the data available for the city of Paris, we set out to provide a multidimensional picture of segregation phenomena, and to

explore segregation indices that can be sensibly and robustly defined from this picture.

Specifically, in this paper, we report on the first steps towards the definition of a data-based segregation index *via* Self-Organizing Maps [29]. Indeed the first part of our work consists in producing a multi-scale description of various variables, from local census blocks to the whole city. A specific novelty of our approach relies on the use of SOM algorithm to incorporate the available data. SOM’s intrinsic mutlidimensionality presents many benefits for the study of such a complex system as a city. This has only been recently noted [30], [31] and our work shall hopefully contribute to SOM becoming a standard tool in urban sociology and geography.

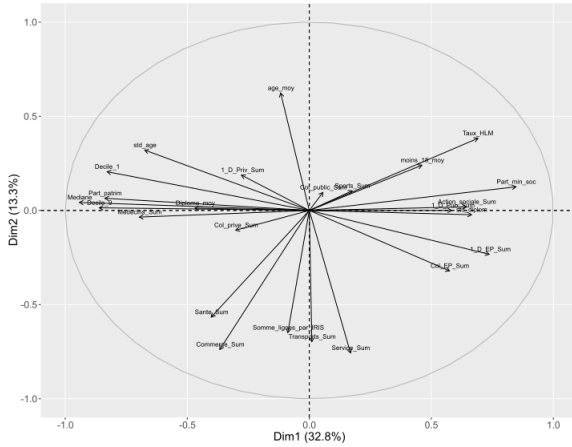
In section II, we describe the pre-processing applied to the available public data, as well as the particular choice of variables we have opted for in this paper. In section III, we examine approaches with some standard statistical tools such as PCA, before turning to SOM in the subsequent section. Finally, we explore a potential segregation index based on the results produced by SOM.

## II. DATA AND VARIABLES

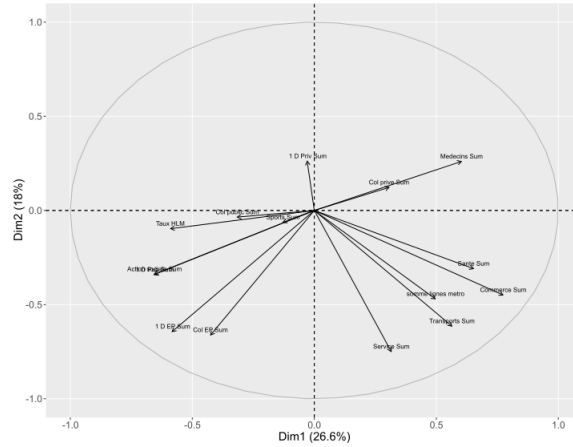
Depending on its source and its type, available data comes in various formats. We used databases from INSEE (*Institut National de la Statistique et des Études Économiques*, France’s national agency for economical data), IGN (*Institut Géographique National*, France’s national agency for geographical data) and RATP (*Régie Autonome des Transports Parisiens*, Paris public transport agency).

### A. Individual data

The individual level is the household one. At this level, INSEE provides data on many characteristics such as the number and age of household members, education level of the household head, size and type of the house, etc. We have aggregated data from this level to the census block level. This allows for instance to compute the fraction of social housing in a given census block.



(a)



(b)

Fig. 1. PCA on all variables (a) and solely on the variables of set 3 (b).

Set	Label	Variable
1	Decile.1	1st decile of income dist.
1	Mediane	Median of income dist.
1	Decile.9	9th decile of income dist.
1	Part_patrim	Share of financial and patrimonial income
1	Part_min_soc	Share of minimum social benefits
2	age_moy	Average age
2	std_age	Standard deviation of age dist.
2	moins_18_moy	Average number of inhabitants under 18 y.o.
2	Diplom_moy	Average level of education (1: pre-secondary; 5: postgrad.)
2	std_diplom	Standard deviation of education level
3	taux_hlm	Fraction of social housing
3	1_D_EP_Sum	Number of EP primary schools
3	1_D_Public_Sum	Number of state primary schools
3	Col_EP_Sum	Number of EP secondary schools
3	Col_Public_Sum	Number of state secondary schools
3	1_D_Priv_Sum	Number of non-state primary schools
3	Col_Priv_Sum	Number of non-state secondary schools
3	Commerce_Sum	Number of shops
3	Services_Sum	Post offices, local administration offices
3	Sports_Sum	Number of sports facilities
3	Action_sociale	Social services offices
3	Medecins_Sum	Number of medical doctors (GPs and specialists)
3	Sante_Sum	Number of hospitals, pharmacies
3	Transport_Sum	Number of train stations and travel agents
3	Somme_lignes	Number of metro and tram stops (within 800m)

TABLE I. Variables used for this exploratory study. (When counting shops, schools, services etc, the 10 nearest neighbouring blocks of a given IRIS are taken into account.)

### B. Census block data

In INSEE data, census blocks are called IRIS (*Ilots Regroupés pour l'Information Statistique*). Unfortunately, neither do they correspond to a fixed surface area nor to a fixed number of inhabitants, although they correspond on average to blocks of around 2,000 inhabitants. Thus, the city of Paris, with a total population of over 2 million people comprises just under 1,000 IRIS. Some IRIS blocks are also purely geographical, with no or very few inhabitants. They appear nonetheless in INSEE data because of services and facilities they may offer. They also appear in the contour data provided by IGN to draw geographical maps and spatial representations.

At the census block level, INSEE provides data such as the number and types of shops, the number and types of public

service offices, the number and type of health facilities. They also provide quantiles of the income distribution within each census block – except those where the number of households is too small and combined with a high level of income. For this and other similar reasons, the number of IRIS for which data is available differs from one variable to the other.

### C. Transport network

The metropolitan authority for transportation provides geographical coordinates for all access points to underground, tramway and bus stations in the Paris area. Using this data (a version of it processed by the OpenStreetMap project [32]), we have been able to compute for each census block the number of underground and tramway lines available within an 800 meter radius (from the centroid of the block). Note that a station with two lines counts twice as a station with just one line.

### D. Variables

For this exploratory case study, we have retained three sets of variables for each census block:

- Set 1 revenue and income: first and ninth deciles as well as median of the income distribution, fraction of revenue coming from assets and other patrimonial sources, fraction of revenue coming from minimal social benefits<sup>1</sup>. From the available data, these variables could be computed for 853 IRIS blocks in Paris.
- Set 2 population: age (average and standard deviation), number of people under 18 years old, education level (coded in 5 groups, average and standard deviation). From the available data, these variables could be computed for 943 IRIS blocks in Paris.
- Set 3 urban facilities and services: rate of social housing, access to public transport, number of shops, access to medical and health services, number of sports facilities, number of primary and secondary schools

<sup>1</sup>These are social benefits paid to prevent people from falling into extreme poverty. They vary from 300 euros to about 800 euros per month.

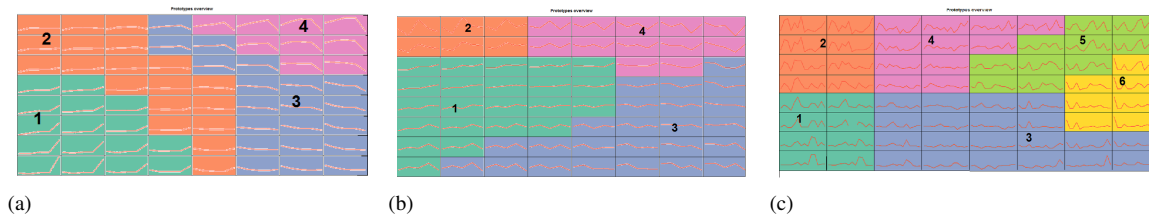


Fig. 2. Kohonen maps for the variables of set 1 (subfigure a), 2 (b) and 3 (c), with values of the variables for the prototypes in each class.

(including primary schools in special urban and education development projects – called *éducation prioritaire* (EP) in France). From the available data, these variables could be computed for 980 IRIS blocks in Paris.

A full list of the variables used in this study is given in Table I.

### III. EXPLORATORY ANALYSIS

Let us first conduct some standard statistical analysis, in particular a Principal Component Analysis (PCA) on our set of variables. As one may expect and as may be seen on Fig 1(a), some variables are clearly correlated or anti-correlated while others seem uncorrelated. For instance, there is a correlation between the fraction of revenue coming from social benefits and the rate of social housing in a given IRIS block. On the contrary, there seems to be very little correlation between the number of shops and services and income levels. Nor, perhaps more surprisingly, between age and income levels.

Such observations can be confirmed and made more precise by looking at the correlation matrix.

Each set of variables may also be analyzed separately, yielding further indications as to correlations. Notably, analysing the third set of variables shows a negative correlation between variables such as the rate of social housing and the number of EP primary schools on the one hand and variables such as the number of shops and access to public transport on the other (see Fig. 1(b)).

If PCA allows to identify such positive and negative correlations, and proves useful to describe a couple of variables, it is clearly ill-suited to a multivariable context, where the mere number of variables leads to poorly-readable and interpretable results. This calls for an intrinsically multidimensional approach, which we introduce in the next section.

### IV. SELF-ORGANIZING MAP APPROACH

We use a multidimensional classification algorithm known as Self-Organizing Map and first introduced by T. Kohonen [29]. All results were obtained using the R-package SOMbrero [33], which performs SOM combined with a hierarchical ascendant classification (HAC). Given that we are dealing with just under one thousand census blocks, we choose to train the algorithm to produce an 8x8 map for each of the three sets of variables, and then to classify individual IRIS blocks into four clusters for sets 1 and 2 and six clusters for the third set (the number of clusters is chosen so as to provide a meaningful classification, with a balance between too little separation among clusters and too few details in their

Cluster	1st decile	Median income	9th decile	Revenue from assets
1	8173	18410	38317	9.5
2	9834	26567	56911	16
3	12542	32305	69360	21
4	13120	43152	124427	39
All	10672	28411	65309	19

TABLE II. Per-cluster averages of some of the variables in set 1 (in euros for quantiles of income distribution, in percent for the share of revenue drawn from financial and other assets).

Cluster	Age	Children per household	Education
1	39.2	0.31	2.7
2	40.4	0.74	2.6
3	44.4	0.32	2.7
4	46.4	0.48	2.2
All	42.6	0.37	2.6

TABLE III. Per-cluster averages of some of the variables in set 2 (education level is from pre-secondary (1) to postgraduate (5) level).

description). The results are shown on Figures 2(a), 2(b), and 2(c), and we proceed to analyze them.

#### A. SOM on set 1

The first set of variables correspond to income and revenue variables. These are the most commonly used in socio-economic segregation studies (along with ethnic group variables – let us recall here that ethnic statistics are not available in France). As can be seen on Figure 2(a), SOM followed by HAC yields here well ordered clusters, parallel to the diagonal, corresponding to four easily identifiable types of census blocks (see Table II). There are blocks (cluster 1) where the population is clearly poorer than the Parisian average, with a first decile of just above 8,000 euros per year, and a share of revenues coming from social benefits almost equal to 1/20. At the other end of the spectrum, blocks where the population is not only richer than average (specifically the top 10 percent are much wealthier, with a ninth decile above 124,000 euros per year) but also with a very substantial part of revenues coming from financial and other patrimonial assets: almost 40 percent on average (cluster 4). In between these two clusters, the other two types of blocks correspond to lower (2) and upper (3) middle classes, with again a difference in the level of patrimonial income (15 percent vs 20 percent). Patrimonial income thus seems to work as a sort of order variable, and is significantly correlated with spatial segregation: Figure 3 shows indeed a high level of spatial homogeneity for the clusters derived from the first set of variables.



Fig. 3. Spatial distribution of clusters 1 to 4 for set 1.



Fig. 4. Spatial distribution of clusters 1 to 4 for set 2.

### B. SOM on set 2

SOM followed by HAC run on the second set of variables yields a less structured clustering than in the case of the first set of variables (see Figure 2(b)). However, one can identify four groups and the following underlying trends: the bottom of the Kohonen map (groups 1 and 3) corresponds to blocks with fewer children and a higher education level than the top part (groups 2 and 4). Also cluster 1 has younger heads of households on average than cluster 3. Similarly in cluster 2 the population is comparatively younger than in cluster 4, and with a higher education level (see Table III).

Spatially, clusters have a wider distributions (Fig. 4), but one still notes that certain areas are particularly representative of a given cluster, eg the northern-north-eastern part of the city for group 4.

### C. SOM on set 3

Processing the third set of variables with SOM and HAC allows one to distinguish six well-identified types of census blocks, that pave the Kohonen map (Fig. 2(c) and Table IV):

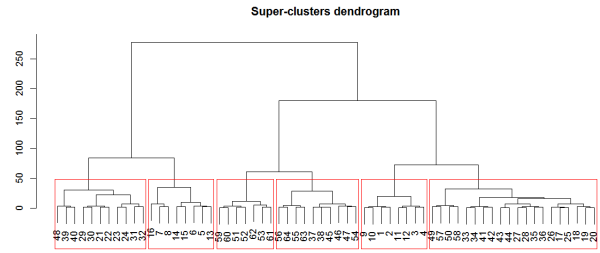


Fig. 5. Dendrogram of the SOM + HAC clusters for the variables of set 3.

- 1) areas with more medical services, more private schools, fewer shops and less access to public transports;
- 2) areas with many shops, facilities and the highest access to public transports;
- 3) areas with a slight concentration of social housing (12 percent) and below average for all other variables;
- 4) areas with a high level of access to public transports, many shops and facilities and also a certain number of EP primary schools;
- 5) areas with a significant proportion of social housing (26 percent), a significant number of EP primary schools, below-average access to public transport
- 6) areas with the highest proportion of social housing (75 percent) but fewer EP schools than the previous group, below-average access to public transport and other facilities.

On this set of variables, a multidimensional approach such as ours sheds light on residential patterns, allowing to refine as one sees fit the level of description. For instance, in this case if one opts for only five clusters, groups number 5 and 6 will be merged, as may be seen from their proximity on Kohonen's map as well as on the dendrogram (Fig. 5).

If one looks at the spatial distribution of the clusters (Fig. 6), there are again some district areas (called *Arrondissements* in Paris) that emerge as particularly representative of a given cluster: 5th, 6th and 7th *Arrondissements* for group 1, 8th for group 2, parts of the 12th, 15th and 16th for group 3, parts of the 1st, 2nd, 3rd and 10th for group 4, 18th, 19th and 20th for group 5, and 13th and the outskirts of the city for group 6.

## V. SEGREGATION INDICES

One of the main difficulties arising in the study of segregation phenomena is that of defining indices to measure the level of segregation [28], [34].

Since most INSEE and other public data in Paris is only available at the IRIS level, let us think at this level. Then measuring segregation amounts to quantifying two things:

- 1) how different IRIS blocks are from one another (that is individual census blocks are not all alike, eg some have a younger population than others, some have a richer population than others);



Cluster	Social housing	Medical doctors	EP schools	Shops	Public transport
1	7	269	0.36	267	6.8
2	3	262	0.70	637	13.1
3	12	160	0.95	203	6.5
4	8	153	3.63	430	12.1
5	26	108	9.26	194	6.3
6	75	122	3.77	138	5.3
All	18	173	2.77	288	8

TABLE IV. Per-cluster averages of some of the variables in set 3 (social housing rate is a percentage per block, other variables are raw numbers for each block and its ten nearest neighbours; access to public transport is the number of lines within 800 meters of a block's centroid).

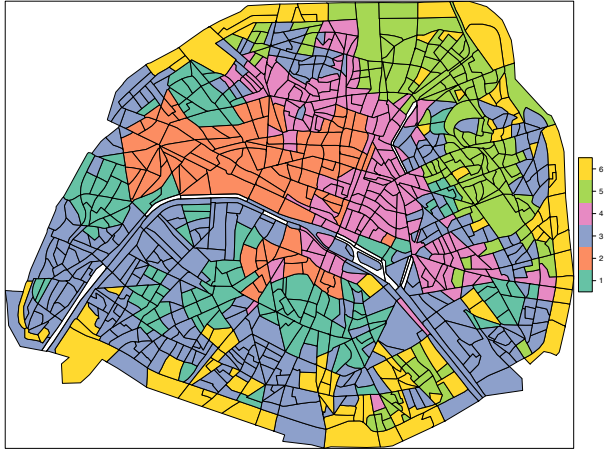


Fig. 6. Spatial distribution of clusters 1 to 6 for set 3.

- 2) how much spatial concentration occurs for IRIS blocks of a given type (in terms of the SOM clusters obtained in the previous section, this means how far one stands from a uniform distribution over the whole city for blocks belonging to each cluster).

The first point may be addressed using the stochastic nature of the SOM algorithm. Indeed, if all blocks were (almost) identical for a given set of variables then the clustering obtained would not be very robust to a re-run of the algorithm with a different seed.

The second point may be treated from two perspectives. One can measure the geographical dispersion of blocks in a given cluster. Or, one can make use of one of SOM's specific properties. Indeed, proximity on the Kohonen map means proximity in the state space of the chosen set of variables. Thus comparing Kohonen distances and geographical distances gives a measure of segregation between blocks from different clusters.

In this paper, we focus on the latter point, and we introduce a new index that will be studied in detail in a forthcoming paper. We also present here results and comparisons using existing indices. Note that a first, simple segregation index that can be computed with any clustering algorithm is the following: for each block, compute the fraction among its  $k$ -nearest neighbours of blocks belonging to the same cluster. (See Index 1 in Table V.)

Set of variables	Index 1	Index 2	Index 3	Index 4
1	0.67	9.2	0.23	0.43
2	0.60	4.85	0.21	0.46
3	0.60	10.41	0.27	0.46

TABLE V. Median values of various segregation indices for the clusters produced by SOM + HAC on our three sets of variables. Index 1 computes for each IRIS block the fraction of blocks belonging to the same cluster among its ten nearest neighbouring blocks. Index 2 is the Kullback-Leibler distance: it measures divergence between the local distribution of clusters (on the 10 nearest neighbours of each IRIS) and the distribution of clusters in the whole city. Index 3 is the  $H$ -index defined in [38], [39] and implemented in the R package seg [40]. Index 4 is the Gini index [41].

#### A. Some entropy and information theory indices

A class of segregation indices are the so-called entropy indices [35], [36]. Based on information theoretic entropy, they measure the difference between the entropy of the global distribution at the city scale and local distributions (we shall consider here each IRIS combined with its 10 nearest neighbouring IRIS). For instance, focusing on the set of variables (income and revenues), an entropy-based segregation index, built on the Kullback-Leibler divergence [37], yields the results shown on figure 7. (See also Index 2 in Table V.)

Another information theory index is the  $H$  index defined in [38], [39], implemented in the R package seg [40]. Particularly suited to multigroup segregation, it may be used here if one arranges for example to describe each IRIS by a 4-vector, the components of which quantify proximity to each of the 4 clusters produced by SOM run on the first set of variables. If all IRIS blocks were the same, the value of  $H$  would be 0 ( $10^{-15}$  on simulated data). With IRIS blocks of mainly two types but well mixed in a checker-board manner on the Parisian grid,  $H$  has a value of 0.12. These values can serve as baselines with which to compare the values obtained with the SOM-clusters on three sets of variables: 0.23 for set 1, 0.21 for set 2, and 0.27 for set 3. (Groups of 10 nearest neighbours have again been used here to pre-process the cluster data in the computation of  $H$ .)

A summary of values obtained for the three standard segregation indices described above, plus the Gini index [41], is given in Table V.

#### B. SOM-based segregation index

As mentioned in the introduction of this section, a new integrated segregation index can be obtained by comparing Kohonen distances and geographical distances for all pairs of IRIS blocks. Indeed, suppose all blocks fall within one of four main clusters and suppose these clusters exhibit a very strong spatial concentration: *eg* cluster 1 is mainly located in the lower



Fig. 7. Entropy-based segregation index (Kullback-Leibler divergence) for the clusters produced by SOM run on set 1. The index measures the divergence between the local distribution of clusters (on the 10 nearest neighbours of each IRIS) and the distribution of clusters in the whole city. The lighter the colour, the smaller the divergence.

left-hand quadrant of the city, cluster 2 mainly in the upper left-hand quadrant, and so on. Then blocks with distance 0 or 1 on the Kohonen map (that is in the same class or in a neighbouring one – recall that in the SOM run here we had 64 classes and then 4 or 6 clusters) will tend to be close to one another on the geographical map too. On the contrary, blocks far apart on the Kohonen map will also tend to be geographically separated. This is indeed what one observes for instance for the variables of set 1, as can be seen on Figures 8 and 9.

A full study of the new segregation index that can be defined in such a way, including comparisons with existing indices on simulated data, will be published in a forthcoming paper.

## VI. CONCLUSION AND PERSPECTIVES

The use of self-organizing maps (combined with HAC) to explore socio-economical and geographical data reveals many promising features. Not only is it intrinsically multidimensional, but it allows for the definition of new segregation indices. Also, the clustering provided is very much suited to interdisciplinary work: for instance, typologies of urban areas inferred from sociological surveys and other forms of studies can be compared to clusters obtained from SOM, and their robustness can be tested by performing a large number of runs with different seeds.

In future work, we will seek to develop a full methodological framework in collaboration with geographers and social scientists. First we will extend further the exploratory study conducted here on the city of Paris to include the full metropolitan area. More variables should be taken into account, bus networks for example, as well as indicators of the service quality and frequency on each line. Comparisons with other cities, both at the European and at the global level will be of particular interest, especially to observe the impact of different public policies.

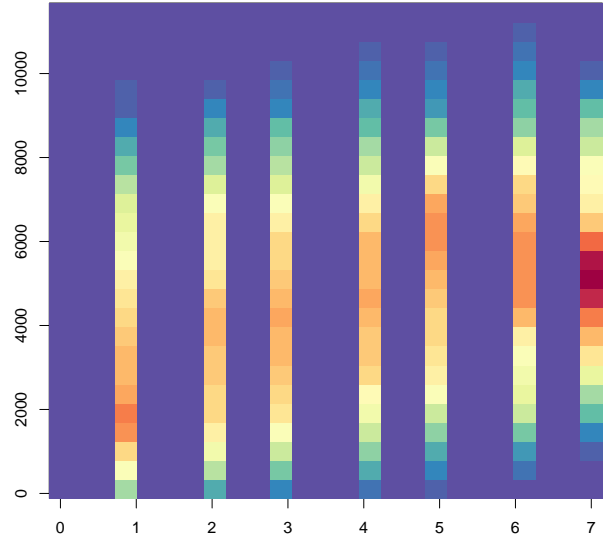


Fig. 8. Visual representation of a SOM-based segregation index for IRIS blocks considered with the variables of set 1. For each value of distance on the 8x8 Kohonen map, the density of pairwise geographical distances is shown (red means higher density). At a point with coordinates (0, 2000) one reads the probability density to find a pair of IRIS blocks that are in the same Kohonen class while their centroids are 2 km apart in the actual city. If IRIS blocks from each Kohonen class were uniformly scattered across the city, geographical distance distributions would be the same for every Kohonen distance.

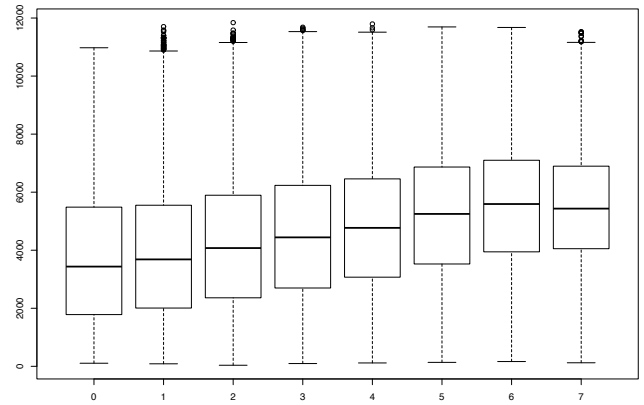


Fig. 9. A box plot representation of the index presented on Figure 8.

## ACKNOWLEDGMENT

The authors would like to thank Cynthia Faure and Marco Corneli for their help and advice on certain aspects of the programming.

## REFERENCES

- [1] T. C. Schelling, "Models of segregation," *The American Economic Review*, vol. 59, no. 2, pp. 488–493, 1969.
- [2] —, "Dynamic models of segregation," *Journal of mathematical sociology*, vol. 1, no. 2, pp. 143–186, 1971.
- [3] W. A. Clark, "Residential preferences and neighborhood racial segregation: A test of the Schelling segregation model," *Demography*, vol. 28, no. 1, pp. 1–19, 1991.
- [4] J. Crane, "The epidemic theory of ghettos and neighborhood effects on dropping out and teenage childbearing," *American journal of Sociology*, vol. 96, no. 5, pp. 1226–1259, 1991.
- [5] M. Pollicott and H. Weiss, "The dynamics of Schelling-type segregation models and a nonlinear graph laplacian variational problem," *Advances in Applied Mathematics*, vol. 27, no. 1, pp. 17–40, 2001.
- [6] M. W. Macy and R. Willer, "From factors to factors: computational sociology and agent-based modeling," *Annual review of sociology*, vol. 28, no. 1, pp. 143–166, 2002.
- [7] A. J. Laurie and N. K. Jaggi, "Role of 'vision' in neighbourhood racial segregation: a variant of the Schelling segregation model," *Urban Studies*, vol. 40, no. 13, pp. 2687–2704, 2003.
- [8] R. Pancs and N. J. Vriend, "Schelling's spatial proximity model of segregation revisited," *Journal of Public Economics*, vol. 91, no. 1, pp. 1–24, 2007.
- [9] W. A. Clark and M. Fossett, "Understanding the social context of the Schelling segregation model," *Proceedings of the National Academy of Sciences*, vol. 105, no. 11, pp. 4109–4114, 2008.
- [10] A. Singh, D. Vainchtein, and H. Weiss, "Schelling's segregation model: parameters, scaling, and aggregation," *Demographic Research*, vol. 21, p. 341, 2009.
- [11] S. Grauwil, E. Bertin, R. Lemoy, and P. Jensen, "Competition between collective and individual dynamics," *Proceedings of the National Academy of Sciences*, vol. 106, no. 49, pp. 20622–20626, 2009.
- [12] A. D. Henry, P. Pralat, and C.-Q. Zhang, "Emergence of segregation in evolving social networks," *Proceedings of the National Academy of Sciences*, vol. 108, no. 21, pp. 8605–8610, 2011.
- [13] A. Banos, "Network effects in Schelling's model of segregation: new evidence from agent-based simulation," *Environment and Planning B: Planning and Design*, vol. 39, no. 2, pp. 393–405, 2012.
- [14] R. Durrett and Y. Zhang, "Exact solution for a metapopulation version of Schellings model," *Proceedings of the National Academy of Sciences*, vol. 111, no. 39, pp. 14036–14041, 2014.
- [15] D. Vinković and A. Kirman, "A physical analogue of the Schelling model," *Proceedings of the National Academy of Sciences*, vol. 103, no. 51, pp. 19261–19265, 2006.
- [16] D. Stauffer and S. Solomon, "Ising, Schelling and self-organising segregation," *The European Physical Journal B*, vol. 57, no. 4, pp. 473–479, 2007.
- [17] L. Dall'Asta, C. Castellano, and M. Marsili, "Statistical physics of the Schelling model of segregation," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 07, p. L07002, 2008.
- [18] L. Gauvin, J. Vannimenus, and J.-P. Nadal, "Phase diagram of a Schelling segregation model," *The European Physical Journal B*, vol. 70, no. 2, pp. 293–304, 2009.
- [19] C. Castellano, S. Fortunato, and V. Loreto, "Statistical physics of social dynamics," *Reviews of modern physics*, vol. 81, no. 2, p. 591, 2009.
- [20] L. Gauvin, J.-P. Nadal, and J. Vannimenus, "Schelling segregation in an open city: a kinetically constrained blume-emery-griffiths spin-1 system," *Physical Review E*, vol. 81, no. 6, p. 066120, 2010.
- [21] T. Rogers and A. J. McKane, "A unified framework for Schellings model of segregation," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2011, no. 07, p. P07006, 2011.
- [22] A. Hazan and J. Randon-Furling, "A Schelling model with switching agents: decreasing segregation via random allocation and social mobility," *The European Physical Journal B*, vol. 86, no. 10, pp. 1–9, 2013.
- [23] V. Cortez, P. Medina, E. Goles, R. Zarama, S. Rica *et al.*, "Attractors, statistics and fluctuations of the dynamics of the Schellings model for social segregation," *Eur. Phys. J. B*, vol. 88, p. 25, 2015.
- [24] A. Sahasranaman and H. J. Jensen, "Dynamics of transformation from segregation to mixed wealth cities," *PloS one*, vol. 11, no. 11, p. e0166960, 2016.
- [25] I. Benenson, E. Hatna, and E. Or, "From Schelling to spatially explicit modeling of urban ethnic and economic residential dynamics," *Sociological Methods & Research*, vol. 37, no. 4, pp. 463–497, 2009.
- [26] E. Hatna and I. Benenson, "The Schelling model of ethnic residential dynamics: Beyond the integrated-segregated dichotomy of patterns," *Journal of Artificial Societies and Social Simulation*, vol. 15, no. 1, p. 6, 2012.
- [27] —, "Combining segregation and integration: Schelling model dynamics for heterogeneous population," *Journal of Artificial Societies & Social Simulation*, vol. 18, no. 4, 2015.
- [28] S. F. Reardon and D. OSullivan, "Measures of spatial segregation," *Sociological methodology*, vol. 34, no. 1, pp. 121–162, 2004.
- [29] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, vol. 43, no. 1, pp. 59–69, 1982.
- [30] D. Arribas-Bel, P. Nijkamp, and H. Scholten, "Multidimensional urban sprawl in europe: A self-organizing map approach," *Computers, Environment and Urban Systems*, vol. 35, no. 4, pp. 263–275, 2011.
- [31] C. Wei, P. Cabrera-Barona, and T. Blaschke, "Local geographic variation of public services inequality: does the neighborhood scale matter?" *International journal of environmental research and public health*, vol. 13, no. 10, p. 981, 2016.
- [32] (2012) Intégration des stations du réseau ferré RATP. Accessed March 2017. [Online]. Available: <http://openstreetmap.fr/blogs/cquest/stations-ratp>
- [33] J. Boelaert, L. Bendhaiba, M. Olteanu, and N. Villa-Vialaneix, *SOMbrero: An R Package for Numeric and Non-numeric Self-Organizing Maps*. Springer International Publishing, 2014, pp. 219–228.
- [34] M. Batty, "Entropy in spatial aggregation," *Geographical Analysis*, vol. 8, no. 1, pp. 1–21, 1976.
- [35] H. Theil and A. J. Finizza, *A note on the measurement of racial integration of schools by means of informational concepts*. Taylor & Francis, 1971.
- [36] J. Iceland, "The multigroup entropy index," *US Census Bureau*, vol. 31, p. 2006, 2004.
- [37] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.
- [38] S. F. Reardon and G. Firebaugh, "Measures of multigroup segregation," *Sociological methodology*, vol. 32, no. 1, pp. 33–67, 2002.
- [39] S. F. Reardon and D. OSullivan, "Measures of spatial segregation," *Sociological methodology*, vol. 34, no. 1, pp. 121–162, 2004.
- [40] S.-Y. Hong, D. O'Sullivan, and Y. Sadahiro, "Implementing spatial segregation measures in r," *PloS one*, vol. 9, no. 11, p. e113767, 2014.
- [41] C. Gini, "Measurement of inequality of incomes," *The Economic Journal*, vol. 31, no. 121, pp. 124–126, 1921.