

Migration Studies with a Compositional Data Approach: A Case Study of Population Structure in the Capital Region of Denmark

Javier Elío[1(✉)], Marina Georgati[1], Henning S. Hansen[1], and Carsten Keßler[1,2]

[1] Department of Planning, Aalborg University, Copenhagen, Denmark
javierdem@plan.aau.dk

[2] Department of Geodesy, Bochum University of Applied Sciences, Bochum, Germany

Abstract. Computing percentages or proportions for removing the influence of population density has recently gained popularity, as it offers a deep insight into compositional variability. However, data are constrained to a constant sum and therefore are not independent observations, a fundamental limitation for applying standard multivariate statistical tools. Compositional Data (CoDa) techniques address the issue of standard statistical tools being insufficient for the analysis of closed data (i.e., spurious correlations, predictions outside the range, and subcompositional incoherence) but they are not widely used in the field of population geography. Hence, in this article, we present a case study where we analyse at parish level the spatial distribution of Danes, Western migrants and non-Western migrants in the Capital region of Denmark. By applying CoDa techniques, we have been able to identify the spatial population segregation in the area and we have recognised patterns in the distribution of various demographic groups that can be used for interpreting housing prices variations. Our exercise is a basic example of the potentials of CoDa techniques which generate more robust and reliable results than standard statistical procedures in order to interpret the relations among various demographic groups. It can be further generalised to other population datasets with more complex structures.

Keywords: Population geography · Migration · Compositional data · Cluster analysis · House prices

1 Introduction

In population geography, it is often more interesting to analyse proportions, such as the percentage of people in a region with an income below the poverty line or the proportion of people fully vaccinated against COVID-19, than absolute

© The Author(s) 2022
O. Gervasi et al. (Eds.): ICCSA 2022 Workshops, LNCS 13379, pp. 576–593, 2022.
https://doi.org/10.1007/978-3-031-10545-6_39

population numbers. In this sense, the variable of interest is normalised so that it does not depend on the total population of the region [9], thus allowing an intuitive analysis of spatial patterns (e.g., distribution of poverty or vaccinate rates across regions). This normalisation explicitly describes the internal structure of the explored system. However, the data are constrained to a constant sum (e.g., 1 for proportions or 100 for percentages) and, therefore, they are dependent; if the share of one subgroup increases, another one has to decrease to retain the sum [24]. This violates a fundamental assumption of most standard statistical analysis and alternative methods are thus required.

These data are known as compositional data (CoDa) and have been widely used in many kinds of interdisciplinary analysis [2]. Aitchison first described the theoretical background to handle such data based on log-ratio transformations [1]. However, this approach is still not widely used despite the associated issues that arise when analysing them with standard statistical procedures; i.e., spurious correlations, predictions outside the range, or problems with subcompositional coherence [20]. In fact, CoDa methods have been mainly applied in the geosciences but even in this field it is not a standard procedure [5]. Instances of CoDa application range from soil and geochemical surveys [34], water and groundwater studies [6], to the evaluation of the link between indoor radon and topsoil geochemistry [14]. Outside of the geosciences, the technique is gaining popularity across different studies in various fields including the evaluation of urban water distribution [10], health studies [22], nutrition research [7], and forecast of energy consumption structures [32].

Human geography is no exception regarding the limited use of CoDa techniques. Often geographers apply standard statistical and geostatistical tools to analyse compositional data (e.g. percentage of young, working age and elderly population in a region; percentage of rented/owner households; unemployment rates) even though these tools have been designed for unconstrained data and are deemed insufficient or unsuitable for such analysis. Lloyd, among others [19], warned about these problems and introduced tools for dealing with compositional data in population studies. Nonetheless, the research community has not widely adopted these tools with the exception of some recent studies. Specifically, CoDa techniques have been utilised for evaluating socio-spatial segregation [8], studying child mortality levels and trends [13], forecasting population age structure [33], or visualising three part compositions in demographic analysis [27]. To the best of our knowledge, there are no studies that use the full range of CoDa techniques to analyse migration data and only Nowok [23] has proposed the use of ternary diagrams for evaluating migration flows.

In this article, we therefore show the applicability of compositional data techniques in population geography analysing the spatial population structure of the capital region of Denmark as a case study. Our aim is to stress the need of CoDa techniques in this type of studies to get robust and reliable results about the compositional variability of the population. We used parish-level data for the year 2020 from Statistics Denmark to analyse the spatial distribution of the three main population categories as defined by the national statistics based

on migration background: people of Danish origin, Western and non-Western migrants. After performing a log-ratio transformation (i.e., balances), we carried out a hierarchical cluster analysis for detecting areas where migrants settle down preferentially in the Capital region of Denmark. Furthermore, we explored the association between migration, considering also the area of origin, and housing prices.

It is well known that house prices and migration are closely related with profound implications on urban planning. There is a two-way causal relationship between migration and house prices [18]. On the one hand, a rise in house prices increases a household's housing equity and, therefore, ability to migrate, since homeowners have a higher financial flexibility for purchasing a new house. At the same time, high house prices make the house unaffordable, thus limiting the number of potential buyers. This way, price differences between regions where migrants live and regions where they intend to move affect in- and out- migration rates. Moreover, the expectation of future house prices also plays an important role in the decision to move [25]. On the other hand, migration increases housing demand and consequently prices [31]. An example of this effect has been found in Sweden where "a 1% increase in the foreign-born population results in a 0.8% increase in house prices, which increases to 1.2% if internal migration is also accounted for" [30]. However, data that contain relative information (e.g., percentage of young, working age and elderly population; share of the population with a certain education level; or percentage of migrants) were used in the models and thus the analysis would be benefited from applying CoDa techniques to avoid possible issues with non-independent observations.

2 Theoretical Background

The main idea in Aitchison's proposal for analysing compositional data was to transform them in a way that allows their analysis with standard statistical tools, designed for unconstrained data. He therefore introduced the concept of log-ratio transformations: the additive log-ratio transformation (alr) and the centred log-ratio transformation (clr). In 2003, Egozcue [11] proposed a new family of transformations called isometric log-ratio transformations (ilr) to overcome some of the limitations of the alr and clr transformations. However there is no single best transformation and all of them have their strengths and limitations [22]. In all types of transformations special attention should be put on the case of zeros, since the logarithm of 0 is undefined [20,21].

The alr transformation for a compositional vector $x = [x_1, x_2, \ldots, x_D]$, with positive components ($x_i > 0$) summing to a constant ($\sum_{i=1}^{D} = k$), is defined as [20]:

$$alr(x) = log(\frac{x_1}{x_D}, \frac{x_2}{x_D}, ..., \frac{x_{D-1}}{x_D}) \tag{1}$$

The alr transformation is useful for parametric modelling. However, it is not invariant under permutation of the components and it is not isometric between the simplex (the sample space of compositional data; $x = [x_1, ..., x_D] \in S^D$)

and the real space (R^{D-1}) [2,5]. In order to address the limitations of the alr transformation, the clr transformation was proposed:

$$clr(x) = log(\frac{x_1}{g(x)}, \frac{x_2}{g(x)}, ..., \frac{x_D}{g(x)}) \tag{2}$$

where g(x) is the geometric mean of the part of the composition. The clr transformation solves the problem of symmetry and, unlike the alr-transformed variables, ordinary distances can be computed [20]. The clr transformation is useful for generating biplots [3,20], but it cannot be used for parametric modelling [5]. Considering this constraint, Egozcue et al. [11] finally proposed the ilr transformation:

$$y = ilr(x) = (y_1, y_2, ..., y_{d-1}) \in \mathbb{R} \tag{3}$$

where:

$$y_i = \frac{1}{\sqrt{i(i-1)}} \ln \left(\frac{\prod_{j=1}^{i} x_j}{(x_{i+1})^i} \right) for\ i = 1, ..., D-1 \tag{4}$$

The ilr transformation allows using all the standard multivariate procedures [11] but the ilr coordinates may be difficult to interpret. The Sequential Binary Partitions method was therefore developed [12]. The result is a particular case of ilr coordinates (i.e., balances) that represent the relationship between two groups of parts allowing the interpretation of their inner connections. The difficulty is to select the correct partitions for obtaining meaningful interpretations and it should be done based on expert knowledge and/or by compositional biplots [20]. The general formula for balances is:

$$b_i = \sqrt{\frac{rs}{r+s}} \ln \left(\frac{(\prod_{+} x_j)^{\frac{1}{r}}}{(\prod_{-} x_k)^{\frac{1}{s}}} \right) for\ i = 1, ..., D-1 \tag{5}$$

where $\prod_{+}$ and $\prod_{-}$ are the parts coded as + or – in the partitioning scheme and r and s the number of components in the + and – partition.

3 Data and Methods

This section introduces the population data and the data on housing prices used for the study. Moreover, it describes the application of CoDa techniques on the data.

3.1 Population Data at Parish Level

Data at parish level have been obtained from Statistics Denmark [29]. The table contains information about the population at the first day of the year and divides it in five ancestry groups: persons of Danish origin, immigrants from Western

countries, immigrants from non-Western countries, descendants from Western countries, and descendants from non-Western countries. The concepts of '*immigrants and descendants*' and '*western and non-western countries*' do not occur in other countries and are defined by Statistics Denmark (DST). According to DST, western countries are all 28 EU countries, Andorra, Iceland, Liechtenstein, Monaco, Norway, San Marino, Switzerland, Vatican State, Canada, USA, Australia, and New Zealand, while all other countries are non-western countries.

We selected only the data referring on the capital region of Denmark in 2020. We also assumed that immigrants and their descendants behave similarly and thus we merged them to simplify the interpretability of our case study. Finally, we closed the dataset to represent percentages over the total population in each parish. Table 1 and Fig. 1 show the summary statistics of the percentages and the spatial distribution respectively.

Table 1. Summary statistics of population data (in percentage) by parish (N = 127).

Value	Danes	Non-western	Western
Mean	77.8	14.5	7.7
Median	79.6	11.5	7.1
IQR	72.9–84.3	7.7–17.7	4.9–9.7
Range	21.2–94.6	2.7–69.9	2.7–19.9

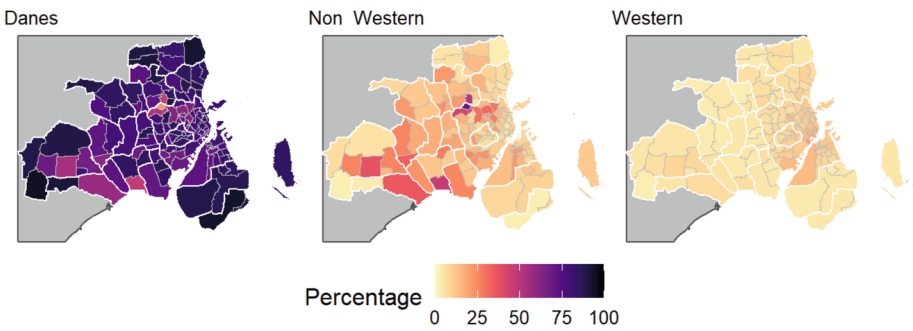

Fig. 1. Population distribution [%] in the capital region of Denmark.

3.2 House Prices

We obtained the individual house prices from the Building and Dwelling Register (BBR - https://teknik.bbr.dk/forside). We used all residences for year-round living (i.e., excluding summer houses and similar seasonal housing) and we selected from the main residential buildings only those that are on the ordinary free trade (sales between parties who are not members of the same family and sales that are not considered as a partial gift) or public sales, assuming that they also

represent a market value. Furthermore, we filtered out dwellings that are not used for residential purposes, are smaller than 10 m^2, or have no value. Colleges and residential buildings for institutions (i.e., different kinds of dormitories) were excluded from the data analysis since they are mainly outside of the free market. We calculated the mean and the median prices in 1.000 Danish kroner per square meter ($kDKK \cdot m^{-2}$) along with the number of dwellings per parish (Table 2 and Fig. 2).

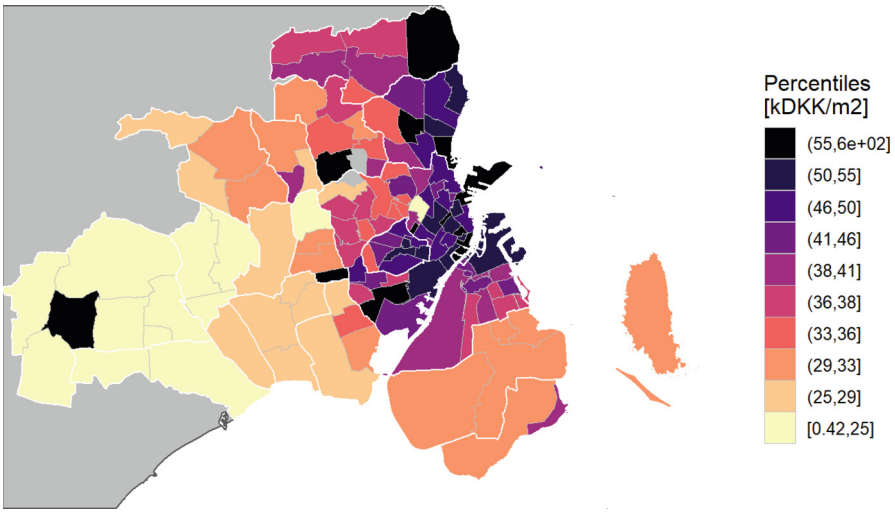

Fig. 2. Spatial distribution of parish-level median house price in 2020.

Table 2. Summary statistics at parish level (N = 125)

Value	N. House	House prices ($kDKK \cdot m^{-2}$).	
		Mean	Median
Mean	150.1	82.8	46.2
Median	127.0	42.3	38.0
IQR	85.0–182.0	33.2– 61.2	31.1–47.4
Range	7.0–637.0	22.1–1,132.9	0.4–602.6

3.3 Compositional Data

The Sequential Binary Partition for our balance calculation has been carried out based on the compositional biplot as presented in Fig. 3(A). The first component differentiates mainly non-Western migrants from other inhabitants, and

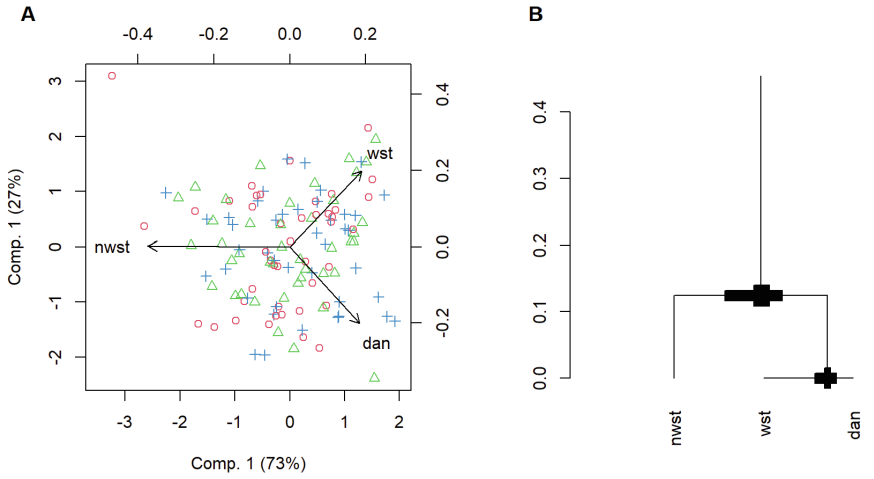

Fig. 3. A) Compositional Biplot, and B) balance-dendrogram of the selected partition.

Table 3. Partition scheme.

Danes	Western	Non-Western	Balance	R	S
1	1	–1	b_1	2	1
1	–1	0	b_2	1	1

it accounts for about 73% of the variance. The second component explains the remaining 27% and mainly separates the native population from Western migrants (Fig. 3A).

Table 3 and Fig. 3B display the selected partition for the balances.

The equations for estimating the two balances are:

$$b_1 = \sqrt{\frac{2}{3}} \ln \left(\frac{(Danes \cdot Western)^{0.5}}{NonWestern} \right) \tag{6}$$

$$b_2 = \sqrt{\frac{1}{2}} \ln \left(\frac{Danes}{Western} \right) \tag{7}$$

where Danes, Western, and Non-Western are the percentages of each population group in the parish.

Using these balances, we performed a hierarchical cluster analysis to investigate whether there are parishes with similar population distributions and whether they are spatially aggregated or not. We used an agglomerative clustering with the Ward's method, which minimises the total within-cluster variance. The analysis was carried out with the function hclust (see complementary material) of the R-software [26]. Finally, we evaluated the spatial autocorrelation of the balances.

4 Results

This section summarises the obtained findings in our case study area. Our aim is to demonstrate the applicability of CoDa techniques not only in migration studies but generally in population geography showing how log-ratio transformations can be used in spatial data analysis (e.g., cluster analysis). For this purpose, we analyse the structure of the population divided in three groups following initially a ternary-balance scheme technique and then examining the similarities among nearby observations and clusters.

4.1 Ternary Diagram of Population Structure

Population structure varies significantly across the capital region of Denmark. The use of colour composition with the rainbow-like surfaces consists an efficient way to visualise the proportions among the three investigated groups and immediately indicates the composition of the population. Figure 4 illustrates the population structure in the ternary plot centred over the compositional mean (80.1%, 7.4%, 12.5% of Danes, Western, and non-Western population respectively). Specifically, the shades of brown indicate parishes with higher proportions of Danes than the compositional mean while the shades of green and pink indicate higher proportions of Western and non-Western migrants correspondingly. As the map shows, Western populations prevail in parishes in the city centre while non-Western citizens tend to settle down in the western peripheral parishes with percentages up to 41.6%, 49.61%, and 69.85% for Husumvold, Haralds and Tingbjerg parishes respectively.

4.2 Balances

The compositional biplot (Fig. 3A) shows similar patterns as observed in the ternary diagram. Non-western migrants dominate component 1, with an opposite direction than Danes and Western migrants. Therefore, we selected a partitioning scheme (Table 3 and Fig. 3) that separates mainly non-Western migrants from Danes and Western migrants (b1), and then Western migrants and Danes (b2). High values of b1 indicate a smaller proportion of non-Western population and high values of b2 a smaller proportion of Western citizens (Fig. 5).

The balances show positive spatial autocorrelation (Table 4), suggesting some degree of spatial structure in the population by its origin. The local indicators of spatial association (LISA - [4] for both balances (Fig. 6) confirm that non-Western migrants tend to live in the peripheral western parishes (blue colours in b1) while Western migrants tend to settle down around the city centre (blue

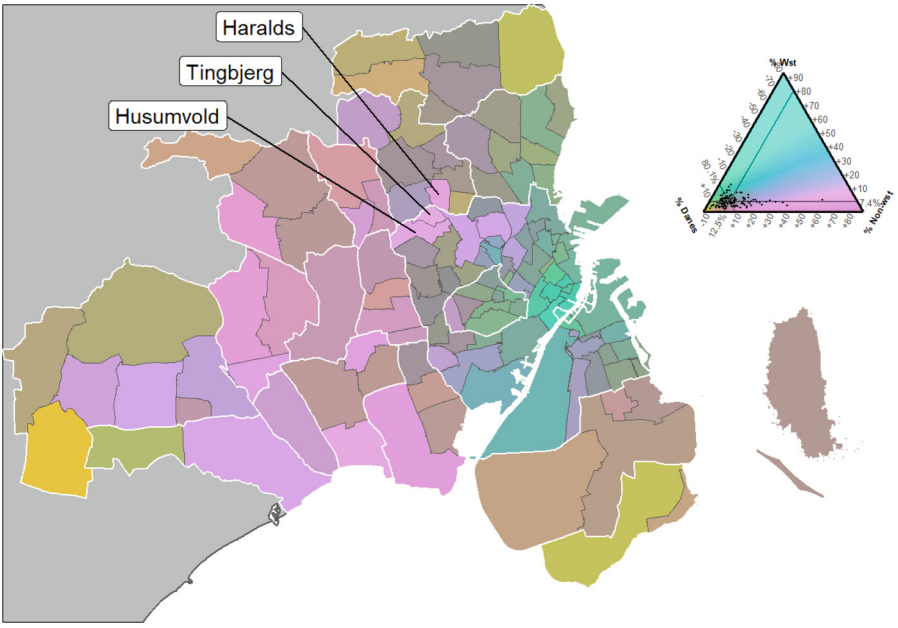

Fig. 4. Ternary diagram of the population distribution in 2020 (Danes - people of Danish origin, Wst - Western population, Non-wst - Non-Western population).

Table 4. Moran's I for each balance.

Balance	Index	Expectation	Variance	Statistic	p-value
b_1	0.470	–0.008	0.003	8.563	5.503×10^{-18}
b_2	0.542	–0.008	0.003	9.837	3.915×10^{-23}

Method: Moran I test under randomisation. Alternative: greater

colours in b2). Furthermore, the presence of non-Western migrants is reduced in the Eastern coast of the capital region (red colours in b1). On the other hand, Danes avoid the city centre and the parishes to the south, north and west tend to have high percentage of national residents (red colours in b2).

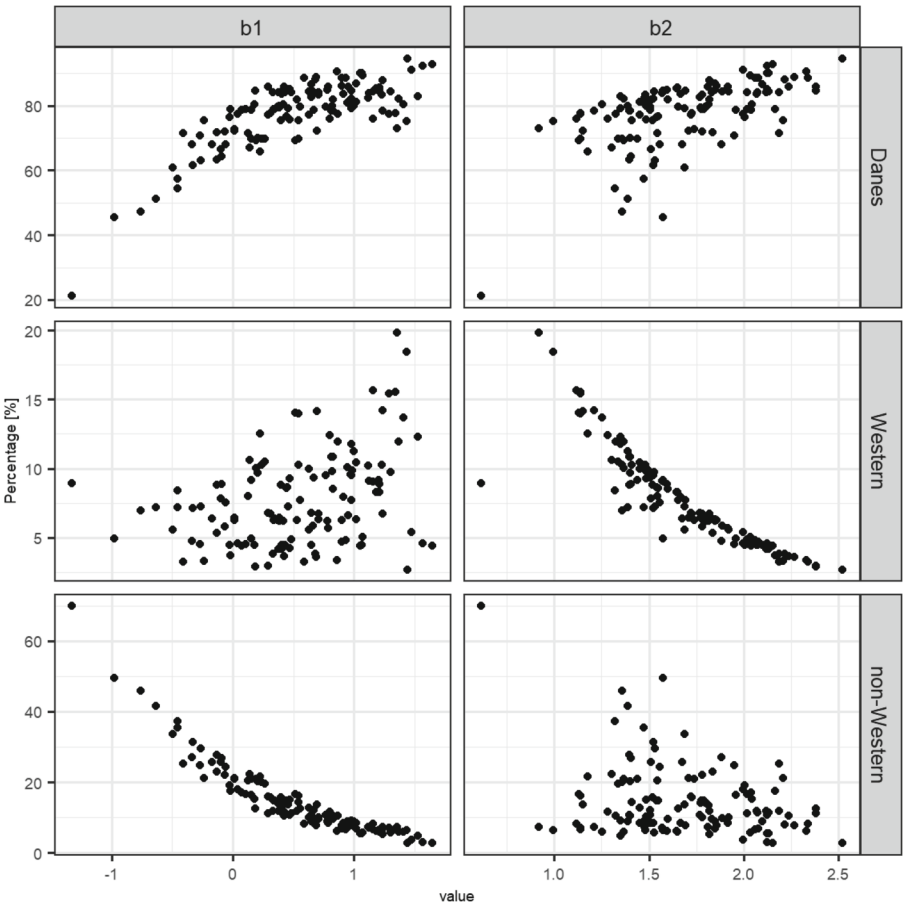

Fig. 5. Balances vs. population distribution percentages.

4.3 Hierarchical Cluster

We identify two main clusters in the data with different proportions of Western migrants and then each one of them is further divided based on the proportion of non-Western migrants. The cluster dendrogram of Fig. 7 shows the first level of division by low and high proportions of Western migrants in the blue boxes (upper) and the second level of division based on the proportions (low/high) of non-Western migrants in the orange boxes (lower). We summarise the compositional means of these four clusters in Table 5 where CL1 and CL2 concentrate a high percentage of Western migrants with a respectively low and high percentage of non-Western migrants. Correspondingly, CL3 and CL4 have low concentration of Western migrants with a respectively low and high concentration of non-Western migrants.

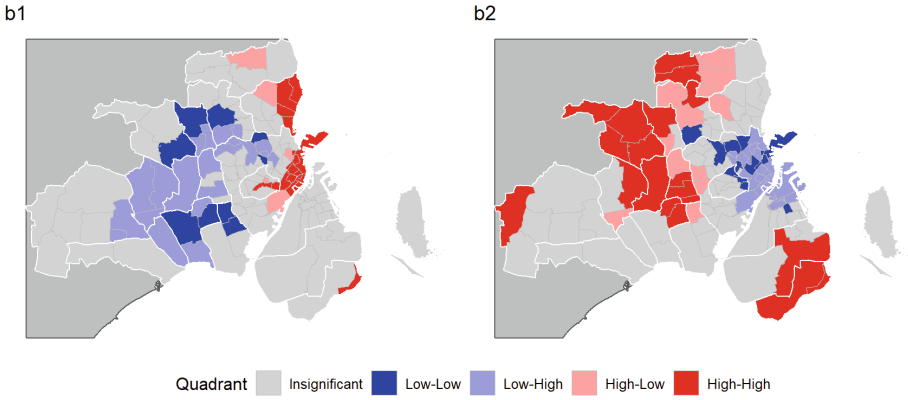

Fig. 6. LISA plots of the two balances (b1 and b2). (Color figure online)

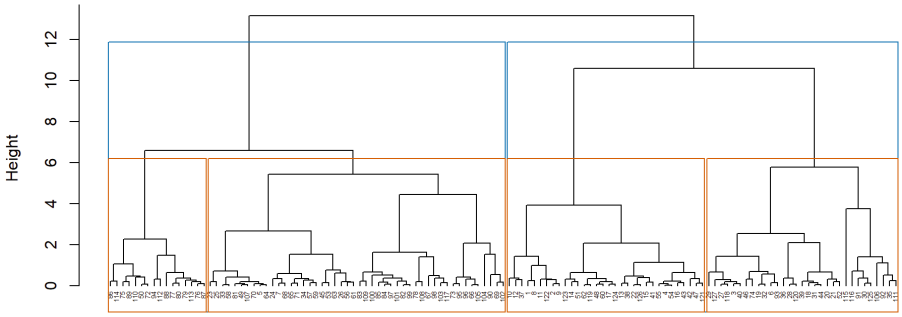

Fig. 7. Cluster dendrogram with the two balances.

Figure 8 spatially maps the distribution of these clusters by parish. Cluster CL1-2 (blue colours) has a median percentage Western migrants of around 10%, while their proportion in CL3-4 (orange colours in Fig. 8) is approximately 5%. This supports the findings of the previous subsections about the tendency of the Western migrants to live in the central parishes. These two main clusters are further divided into four clusters (CL1-2 into CL1 and CL2; and CL3-4 into CL3 and CL4) based on the proportions of the non-Western population. In this regard, CL2 has a higher proportion of non-Western population than CL1 (i.e. 24.1% and 7.8%, respectively), while CL4 has a higher proportion of non-Western than CL3 (i.e. 20.5% and 8.9%, respectively). Again, we observe the preference of non-Western population for the peripheral western parishes (CL2 - light blue; and CL4 - light orange; Fig. 8). Finally, CL3 shows the parishes with the highest proportion of national citizens with values around 85.9%.

Table 5. Compositional mean of the four clusters.

Cluster	N Parishes	Danes	Western	Non-western
1	32	81.0	11.2	7.8
2	31	66.6	9.3	24.1
3	48	85.9	5.2	8.9
4	16	74.8	4.7	20.5

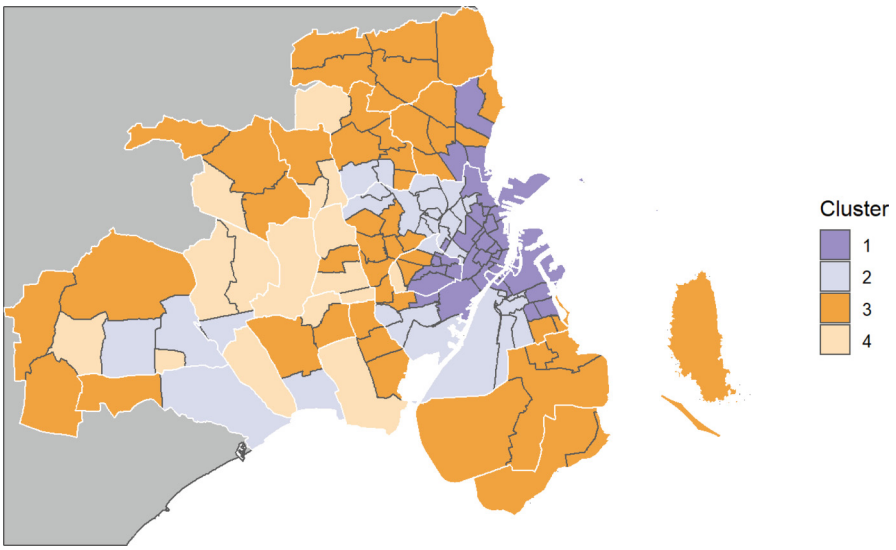

Fig. 8. Spatial distribution of the four clusters.

4.4 House Prices

Analysing the housing prices in the four clusters, there are some clear differences in the mean and median values (Table 6). In general, CL1 and CL2 have higher values (i.e. mean around 57.0 and 41.7 kDKK/m^2; and median of 50.7 and 38.6 kDKK/m^2, respectively) than CL3 and CL4 (i.e. means of 38 and 27 kDKK/m^2, and medians of 35.6 and 27.0 kDKK/m^2, respectively).

A ternary diagram with the median housing prices and the population distribution in Fig. 9 also shows the differences in the median house prices by parish and their association with the clusters. The parishes of CL1, where the proportion of Western migrants is the highest and the proportion of non-Western migrants is relatively low, have the highest median housing prices. Furthermore, house prices decrease with the proportion of non-western migrants, being CL2 the cluster where we observe it clearer. In this sense, in CL2 the proportion of non-Western migrants change remarkably from around 10% to up to approximately 45% (there are two parishes with even more proportion of non-Western migrants; i.e. Haralds and Tingbjerg with 49.61%, and 69.85% respectively, but

Table 6. House price statistics in each cluster.

Cluster	CL1	CL2	CL3	CL4
	(N = 32)	(N = 29)	(N = 48)	(N = 16)
Mean values ($kDKK \cdot m^{-2}$)				
Mean	119.8	82.9	65.7	59.9
Median	57.0	41.7	38.2	27.2
IQR	50.7–98.7	37.3–60.8	31.4–49.8	24.8–33.5
Range	39.4–1,132.9	22.1–554.6	22.2–554.7	22.5–333.6
Median values ($kDKK \cdot m^{-2}$)				
Mean	67.6	37.3	35.8	50.9
Median	50.7	38.6	35.6	27.0
IQR	46.9–54.5	34.2–42.0	30.1–40.2	24.7–32.0
Range	37.1–602.6	0.4–70.8	20.5–55.6	5.1–367.3

there were no data of housing sales on the ordinary free trade or public sales in these parishes in 2020), and the median house price decrease from values around 45 kDKK/m^2 to 25 kDKK/m^2, respectively.

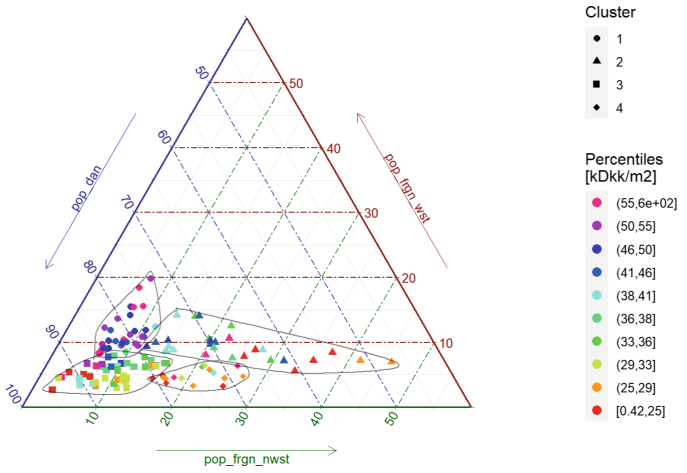

Fig. 9. Population distribution (in percentage) and housing prices (median values in $kDKK \cdot m^{-2}$).

5 Discussion

The hierarchical clusters analysis over the log-ratio transformed data has allowed us to detect four main spatial clusters which clearly characterise parishes according to their population structure (i.e., people of Danish origin, Western migrants,

and non-Western migrants). As expected, Danes are the main population in the region. They tend to avoid the city centre and are more attracted by the parishes of southern Amager and the north-west capital region. Western migrants, on the other hand, prefer the central areas while the proportion of non-Western migrants increases in the western peripheral parishes. Additionally, the ternary diagram (Fig. 4) has also allowed us to graphically identify parishes with a very high percentage of non-Western migrants; i.e., Husumvold, Haralds and Tingbjerg parishes with values up to 41.6%, 49.61%, and 69.85%; respectively, and manly due to a diminution of Danes (51.20%, 45.47%, 21.24%; respectively) rather than Western migrants (7.20%, 4.91%, 8.90%; respectively). These parishes are examples of what the Danish authorities call '*parallel societies*' or '*ghettos*', which trigger political actions [17,28]. Our observations also are in agreement with previous studies [15] and confirm a degree of socio-spatial segregation in the capital region of Denmark.

It is important to note that different phenomena can lead to the same proportions in the data. Compositions only give information about the relative magnitude of its components but not the absolute values [2] and additional information would be needed in order to make inferences with the absolute values. In our study, for example, we have seen the spatial segregation of the population by its origin but we cannot interpret its causes; e.g., if it is an effect of socio-economic segregation or diaspora, where immigrants tend to settle down in areas with existing migrant networks and an ethnicity background similar to their country of origin. This is actually a limitation of any observational study, which helps to make hypothesis about the phenomena we are investigating but further studies would be needed to verify them. CoDa techniques are however more robust than ordinary statistics methods because they alleviate issues with spurious correlations and they avoid problems with sub-compositions since the results obtained by using the whole dataset do not contradict results obtained by only a subset [2].

Regarding house prices and migration, we interpret that the differences in the median house prices we observed in each cluster are probably influenced or related more by their location than the population structure. On the one hand, CL1 and CL2 parishes are central parishes close to the city centre where we can expect more amenities and, therefore, people may be willing to pay more for their house in these areas than for further away ones. On the other hand, the central area of the capital region is more densely populated than the periphery [15] so we can expect a higher demand for houses in these parishes which may lead to an increase in the housing prices. However, if we compare the values within the two main clusters (i.e., CL1 vs. CL2 and CL3 vs. CL4), the prices are lower in the parishes where the proportion of non-Western population is relatively high (around 20%, Table 5). It seems therefore that the relative numbers of non-Western and Western migrants are also associated with the median housing prices (Table 6) and thus population structure should also be taken into account when we evaluate house prices.

While our analysis is limited to a small number of socio-economic variables, our results point to two different kinds of segregation that can be observed in the Danish capital region. The clustering of native Danes and Western immigrants on the one hand and high concentrations of non-Western immigrants in a relatively small number of parishes on the other hand indicates racial segregation. Taking into account house prices, we further see that Western migrants are concentrated in the central parishes, which are characterised by higher prices, whereas the parishes with high numbers of non-Western show lower house prices. This correlation therefore also indicates socio-economic segregation.

6 Conclusions

CoDa techniques are more robust and appropriate than standard statistical and geostatistical methods when we are analysing closed data (e.g. percentages) because they avoid spurious correlations, predictions outside the range, and have no problems with sub-compositional coherence. However, they still are not widely used in population geography. In the present study, we demonstrate a showcase applying CoDa techniques in order to promote their use and adoption in a wider range of applications in population studies and spatial analysis. We specifically carried out a case study from the capital region of Copenhagen and evaluated the obtained results and the ways of interpreting them along with the applicability of the methodology in this field. In this regard, we studied the population distribution divided into people of Danish origin, Western immigrants, and non-Western immigrants and its possible relationship with house prices.

Our analysis facilitates the analysis and the interpretation of the socio-spatial segregation in the capital region of Denmark. We detected four main cluster regions with clear differences in the composition of population in terms of migration backgrounds. Furthermore, we performed an analysis to relate these variation in migration patterns to the median house prices at parish level. CoDa techniques show great potential in recognising such trends and patterns but we only shortly discussed the associations between migration and house prices. Differently put, despite a broad range of factors (e.g. property size, condition, proximity to transportation) influences house prices, they all are considered out of the scope of this study apart from the migration component. Our results, therefore, need to be taken with caution and further investigation is required to evaluate the causal relationship between migration and house prices in the capital region of Denmark.

We showed how balances can be used for alleviating the issue of data interpretation with CoDa methods. There is still some complexity in the interpretation of models based on balances and some researchers [16] have proposed the use of amalgamated logratios (i.e., sum of compositional parts) which produce more comprehensible results with minimal cost. We aim on further investigating their use in population geography in the future. Overall, our exercise is a good example of how CoDa methods could be used for exploratory spatial data analysis of various demographic groups. Although it may be a basic case study with only

three components, it can be generalised to other population datasets with numerous possible applications and advantages ranging from getting a general insight into the compositional variability of population structures to the identification of clusters and the interpretation of complex socio-economic phenomena.

Acknowledgements. This work has been supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No 870649, the project Future Migration Scenarios for Europe (FUME; https://futuremigration.eu); and by the Aalborg University Strategic Fund, through the project: "Global flows of migrants and their impact on north European welfare states - FLOW" (www.flow.aau.dk).

References

1. Aitchison, J.: A new approach to null correlations of proportions. J. Int. Assoc. Math. Geol. **13**(2), 175–189 (1981). https://doi.org/10.1007/BF01031393
2. Aitchison, J.: A concise guide to compositional data analysis. CDA Workshop Girona **24**, 73–81 (2002). https://doi.org/10.2307/4355794
3. Aitchison, J., Greenacre, M.: Biplots of compositional data. J. Roy. Stat. Soc. Ser. C: Appl. Stat. **51**(4), 375–392 (2002). https://doi.org/10.1111/1467-9876.00275
4. Anselin, L.: Local indicators of spatial association-LISA. Geograph. Anal. **27**(2), 93–115 (2010). https://doi.org/10.1111/j.1538-4632.1995.tb00338.x, https://onlinelibrary.wiley.com/doi/10.1111/j.1538-4632.1995.tb00338.x
5. Buccianti, A., Grunsky, E.: Compositional data analysis in geochemistry: are we sure to see what really occurs during natural processes? J. Geochem. Explor. **141**, 1–5 (2014). https://doi.org/10.1016/j.gexplo.2014.03.022
6. Buccianti, A.: Water chemistry: are new challenges possible from CoDa (compositional data analysis) point of view? In: Daya Sagar, B.S., Cheng, Q., Agterberg, F. (eds.) Handbook of Mathematical Geosciences, pp. 299–311. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-78999-6_16
7. Corrêa Leite, M.L.: Compositional data analysis as an alternative paradigm for nutritional studies. Clin. Nutr. ESPEN **33**, 207–212 (2019). https://doi.org/10.1016/j.clnesp.2019.05.011, https://linkinghub.elsevier.com/retrieve/pii/S2405457718304236
8. Cruz-Sandoval, M., Roca, E., Ortego, M.I.: Compositional data analysis approach in the measurement of social-spatial segregation: towards a sustainable and inclusive city. Sustainability **12**(10), 4293 (2020). https://doi.org/10.3390/su12104293, https://www.mdpi.com/2071-1050/12/10/4293
9. Dailey, G.: Normalizing Census Data Using ArcMap. ArcUser (January-March), 52–53 (2006). www.esri.com
10. Ebrahimi, P., Albanese, S., Esposito, L., Zuzolo, D., Cicchella, D.: Coupling compositional data analysis (CoDA) with hierarchical cluster analysis (HCA) for preliminary understanding of the dynamics of a complex water distribution system: The Naples (South Italy) case study. Environ. Sci. Water Res. Technol. **7**(6), 1060–1077 (2021). https://doi.org/10.1039/d0ew01123a
11. Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G., Barceló-Vidal, C.: Isometric logratio transformations for compositional data analysis. Math. Geol. **35**(3), 279–300 (2003). https://doi.org/10.1023/A:1023818214614

12. Egozcue, J.J., Pawlowsky-Glahn, V.: Groups of parts and their balances in compositional data analysis. Math. Geol. **37**(7), 795–828 (2005). https://doi.org/10.1007/s11004-005-7381-9
13. Ezbakhe, F., Pérez Foguet, A.: Child mortality levels and trends: a new compositional approach. Demographic Res. **43**(43), 1263–1296 (2020). https://doi.org/10.4054/DemRes.2020.43.43, https://www.demographic-research.org/volumes/vol43/43/
14. Ferreira, A., et al.: Indoor radon measurements in south west England explained by topsoil and stream sediment geochemistry, airborne gamma-ray spectroscopy and geology. J. Environ. Radioact. **181**, 152–171 (2018). https://doi.org/10.1016/j.jenvrad.2016.05.007, https://linkinghub.elsevier.com/retrieve/pii/S0265931X16301515
15. Georgati, M., Keßler, C.: Spatially explicit population projections: the case of Copenhagen, Denmark. AGILE: GIScience Ser. **2**, 1–6 (2021). https://doi.org/10.5194/agile-giss-2-28-2021, https://agile-giss.copernicus.org/articles/2/28/2021/
16. Greenacre, M., Grunsky, E., Bacon-Shone, J.: A comparison of isometric and amalgamation logratio balances in compositional data analysis. Comput. Geosci. **148**, 104621 (2021). https://doi.org/10.1016/j.cageo.2020.104621
17. Gulis, G., Safi, M., Linde, D.S.: Rapid health impact assessment of a Danish policy document: One Denmark without parallel societies: no ghettos in 2030. J. Public Health (2020). https://doi.org/10.1007/s10389-020-01375-z
18. Jeanty, P.W., Partridge, M., Irwin, E.: Estimation of a spatial simultaneous equation model of population migration and housing price dynamics. Reg. Sci. Urban Econ. **40**(5), 343–352 (2010). https://doi.org/10.1016/j.regsciurbeco.2010.01.002, https://linkinghub.elsevier.com/retrieve/pii/S0166046210000037
19. Lloyd, C.D.: Analysing population characteristics using geographically weighted principal components analysis: a case study of Northern Ireland in 2001. Comput. Environ. Urban Syst. **34**(5), 389–399 (2010). https://doi.org/10.1016/j.compenvurbsys.2010.02.005
20. Lloyd, C.D., Pawlowsky-Glahn, V., Egozcue, J.J.: Compositional data analysis in population studies. Ann. Assoc. Am. Geographers **102**(6), 1251–1266 (2012). https://doi.org/10.1080/00045608.2011.652855, http://www.tandfonline.com/doi/abs/10.1080/00045608.2011.652855
21. Martín-Fernández, J.A., Barceló-Vidal, C., Pawlowsky-Glahn, V.: Dealing with zeros and missing values in compositional data sets using nonparametric imputation. Math. Geol. **35**(3), 253–278 (2003). https://doi.org/10.1023/A:1023866030544
22. McKinley, J.M., et al.: Investigating the influence of environmental factors on the incidence of renal disease with compositional data analysis using balances. Appl. Comput. Geosci. **6**, 100024 (2020). https://doi.org/10.1016/j.acags.2020.100024
23. Nowok, B.: A visual tool to explore the composition of international migration flows in the EU countries, 1998–2015. Demographic Res. **42**, 763–776 (2020). https://doi.org/10.4054/DemRes.2020.42.27, https://www.demographic-research.org/volumes/vol42/27/
24. Pawlowsky-Glahn, V., Egozcue, J.J.: Exploring compositional data with the CoDa-dendrogram. Aust. J. Stat. **40**(1&2), 103–113 (2011). https://doi.org/10.17713/ajs.v40i1&2.202
25. Peng, C.W., Tsai, I.C.: The long- and short-run influences of housing prices on migration. Cities **93**, 253–262 (2019). https://doi.org/10.1016/j.cities.2019.05.011
26. R Core Team: R: A language and environment for statistical computing (2021). https://www.r-project.org/

27. Schöley, J.: The centered ternary balance scheme: a technique to visualize surfaces of unbalanced three-part compositions. Demographic Res. **44**, 443–458 (2021). https://doi.org/10.4054/DemRes.2021.44.19, https://www.demographic-research.org/volumes/vol44/19/
28. Seemann, A.: The Danish 'ghetto initiatives' and the changing nature of social citizenship, 2004–2018. In: Critical Social Policy, pp. 1–20 (2020). https://doi.org/10.1177/0261018320978504
29. Statistics Denmark: KMSTA001: Population 1. January by parish, ancestry and member of the National Church (2021)
30. Tyrcha, A., Abreu, M.: Migration diversity and house prices - evidence from Sweden. SSRN Electron. J., 1–36 (2019). https://doi.org/10.2139/ssrn.3394234, https://www.ssrn.com/abstract=3394234
31. Wang, X.R., Hui, E.C.M., Sun, J.X.: Population migration, urbanization and housing prices: evidence from the cities in China. Habitat Int. **66**, 49–56 (2017). https://doi.org/10.1016/j.habitatint.2017.05.010
32. Wei, Y., Wang, Z., Wang, H., Li, Y.: Compositional data techniques for forecasting dynamic change in China's energy consumption structure by 2020 and 2030. J. Clean. Prod. **284**, 124702 (2021). https://doi.org/10.1016/j.jclepro.2020.124702
33. Wei, Y., Wang, Z., Wang, H., Li, Y., Jiang, Z.: Predicting population age structures of China, India, and Vietnam by 2030 based on compositional data. PLOS ONE **14**(4), e0212772 (2019). https://doi.org/10.1371/journal.pone.0212772
34. Zheng, C., et al.: Application of compositional data analysis in geochemical exploration for concealed deposits: a case study of Ashele copper-zinc deposit, Xinjiang. China. Appl. Geochem. **130**, 104997 (2021). https://doi.org/10.1016/j.apgeochem.2021.104997

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

