

CLUSTER-BASED DEMOGRAPHIC CLASSIFICATION OF RURAL AND SUBURBAN AREAS – EXAMPLE OF LATVIA

Aleksandrs Dahs, Juris Krumins
(University of Latvia)

EXTENDED ABSTRACT

Introduction

Contemporary goals of regional convergence are difficult to achieve, considering the uneven demographic scenery, competition between administrative-territorial units and limited resources available for targeted regional demographic policy measures. In the context of sustainable regional demographic development it is critically important to have an ability to quickly and effectively evaluate demographic risks and assets associated with specific regions or territories.

Shifting administrative boundaries and vague existing classification (e.g. rural vs urban regions) make this task even more challenging. Furthermore, some may argue that differentiation of various types of countryside is even more important than search for the definition of the rural against the urban, proposing different country-specific types of classification for rural and suburban areas (see Stonawska, K. and Vaishar, A., 2018, Hasek, O., 2020). As a result, researchers and policymakers have a need to categorise territories according to their demographic issues and development potential in order to effectively present the actionable information and develop targeted policy measures. The foremost issue is a lack of a defined methodology and criteria for the classification and assessment of territorial units according to their demographic potential and risks.

In this study we aim to develop, test and evaluate methodology for an efficient demographic classification of rural administrative-territorial units using adaptive non-hierarchical clustering, while using the recently reformed municipalities of Latvia as an example. Data used for this purpose include statistical datasets of Latvian municipalities recently recalculated according to the new administrative-territorial division. Our proposed methodology relies on a robust k-medoids PAM (*Partitioning around Medoids*) clustering algorithm, as it is not reliant on a pre-defined measures of distance or dissimilarity between observations and allows for greater freedom in number of indicators and types of data used.

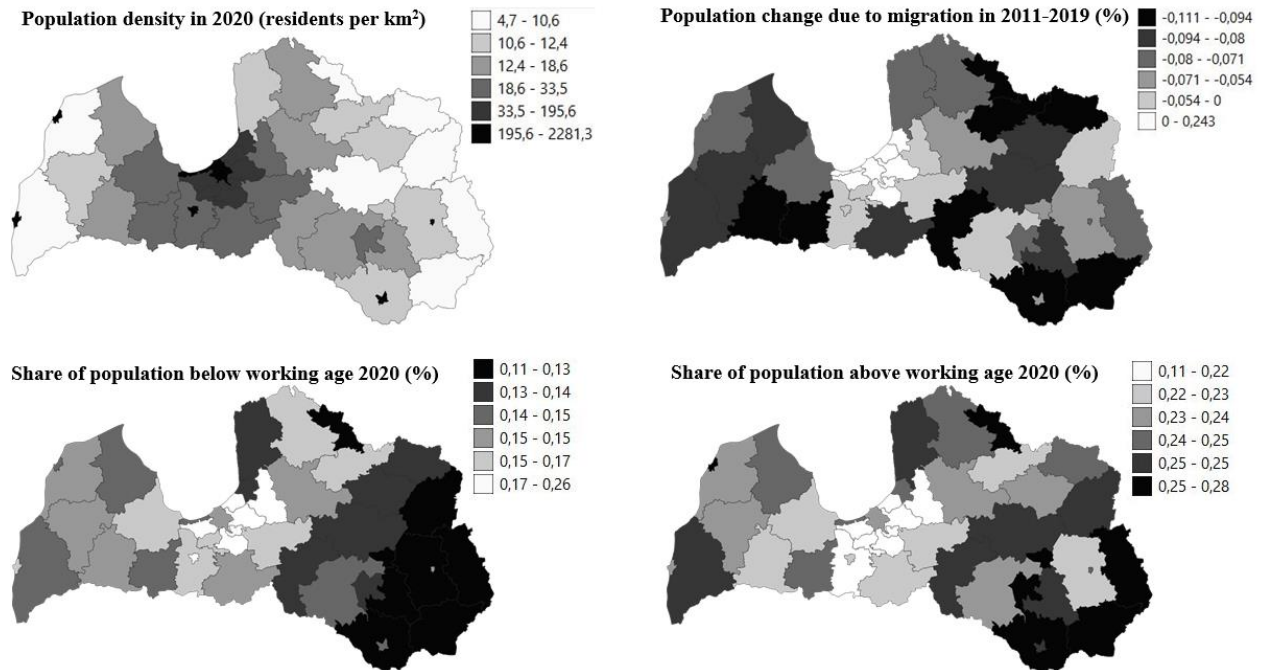
Study results show that with carefully selected data sets, a non-hierarchical cluster analysis employing k-medoids or similar method can become a feasible tool for quick and efficient classification of territorial units according to their socio-economic and demographic characteristics. The established clustering technique provided most useful results when carried out in several iterations, using different data cross-sections. Considering the unique features of national administrative-territorial structures, as well as a variety of demographic challenges and associated local factors, we can also conclude that the demonstrated approach can be most useful, as it permits application of different grouping criteria sets, thus giving room for experimentation and approbation of various theoretical frameworks.

Data and Methodology

In this study we combine processes and methodologies outlined by several previous studies on regional or rural demographic differentiation and focusing on demographic classification of local territorial units.

The administrative territorial reform of 2021 in Latvia (Saeima, 2020) has created a unique testing ground for the potential regional demographic classification methodology, as it changed the shape and structure of the previously established local administrative territories, creating a new municipal structure with a diverse demographic landscape. (Dahs A. et al, 2021). These changes necessitate a better understanding demographic conditions and dynamics in newly formed rural and suburban municipalities in order to elaborate the necessary policy measures.

Although the number of available datasets for the new administrative – territorial structure of Latvia is still quite limited, sufficient data for this study was obtained from the previous works featuring the recalculation of historical demographic data according to the shifting administrative boundaries (Krumins J. et al, 2020) and the national Regional Development Indicator Module (RDIM, 2021) maintained by the State Regional Development Agency. For the purpose of this study we use data for all Latvian municipalities except for the seven biggest State cities (*valstspilsētas*), which are designated as separate administrative units.



Source: authors' elaboration based on RAIM 2021

Figure 1. Examples of key regional demographic indicators for the post-reform municipal structures in Latvia (2011-2020 period)

Considering the characteristics of data sets available in regional demographic scenarios, the non-hierarchical partitioning clustering approach is an obvious first choice, as it is not reliant on a pre-defined measures of distance or dissimilarity between observations and allows for greater freedom in number of indicators and types of data used. Major drawback of this approach – a need to pre-define the number of expected clusters – can be mitigated by comparing the results of various estimations and choosing the best possible fit. Due to the high risk of outlier occurrences in the regional data, we propose using a robust k-medoids PAM (Partitioning around Medoids) algorithm described originally by Kaufman & Rousseeuw (2009).

When determining the number and criteria of possible clusters, there is a limited number of country or region-based studies which can be of use. Keeping in mind that every country and region has unique demographic characteristics, we propose testing several previously elaborated and verified indicators sets.

For example, Joaquin Recano (2017) proposed to group municipalities into three types according to their resilience and risks associated with demographic change. This approach may be adopted to the case of Latvia and data availability using these classification criteria:

1. Resilient spaces - certain demographic stability with population densities higher than the rural average. The impact of emigration is less evident.
2. Spaces of emigration - have a small number of inhabitants and low population density, negative natural growth rates, a relative high level of ageing and major impact of emigration.

3. Spaces at risk of irreversible depopulation - with characteristics at the extreme end of the scale: very small number of inhabitants, low population density and maximum ageing.

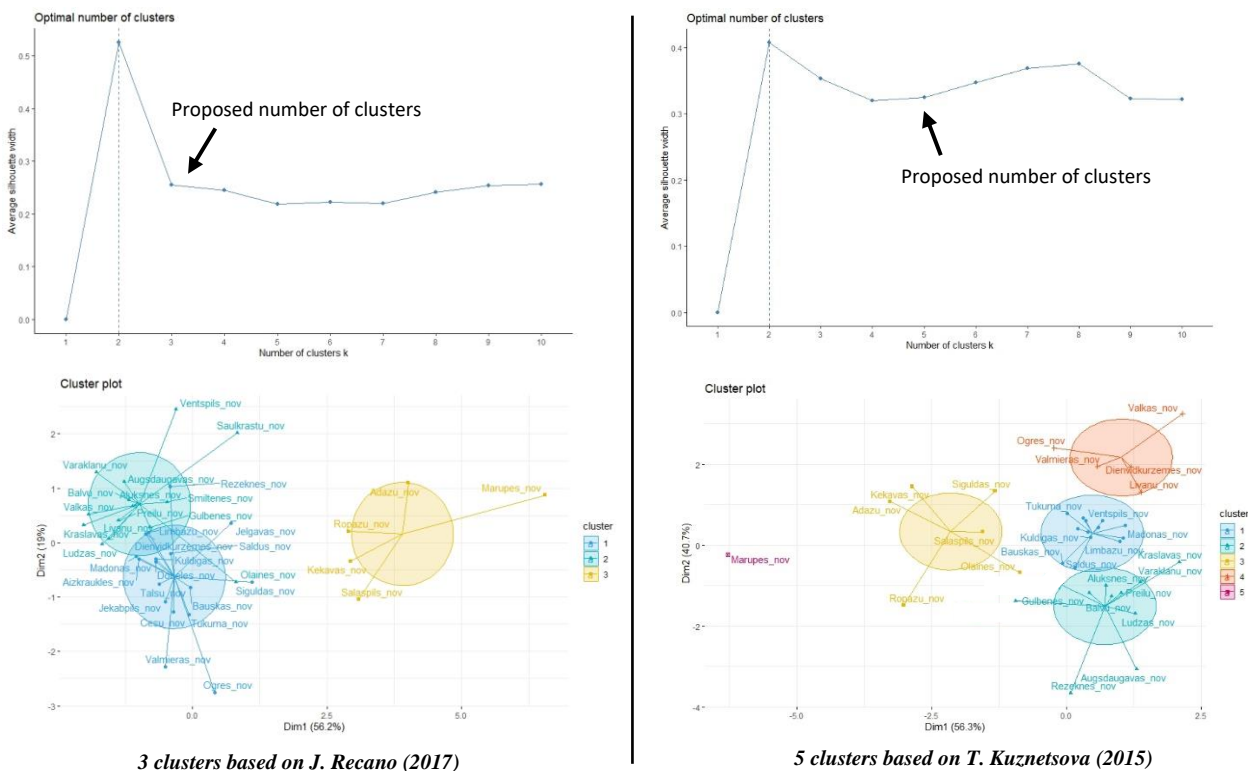
Another noteworthy approach presented by Tatiana Kuznetsova (2015) proposes the use of five clusters based not only on demographic, but also socio economic characteristics of the municipalities under study. With minor adaptations to the available data, categories can be described as follows:

1. Actively growing regions with a relatively favourable population age structure
2. Growing region with significant demographic load of elderly people
3. Declining regions with a relatively favourable age structure of the population
4. Regions with significant demographic load and actively decreasing population
5. Regions with a high load of the elderly population and actively decreasing population

Using the available statistical data and k-medoids PAM clustering algorithm we conduct several clustering estimations, while considering the grouping criteria described above. To estimate the clusters according to each methodology we use only relevant indicators form the statistical tables.

Key findings

To determine the possible number of clusters we use the average silhouette method. The idea is to compute PAM algorithm using different values of clusters and determine the average clusters silhouettes according to the number of clusters. The average silhouette measures the quality of a clustering, with high silhouette width indicating a good clustering (Kaufman and Rousseeuw, 2009).



Source: authors' construction.

Figure 2. Comparison of the clustering results for rural Latvian municipalities using different methodologies and number of clusters

The consecutive re-runs of PAM algorithm on the available data sets clearly indicate that the optimal number of clusters for grouping the newly formed Latvian municipalities according to the clusters based on work of J. Recano (2017) is two. However, a closer look at cluster composition shows that this is caused by a significant disparity between the municipalities in Riga metropolitan area and the rest of the country. With this in mind, we can safely proceed with the proposed three clusters, as it is the next most suitable number indicated by the test. For the clustering based on work by T.

Kuznetsova (2015) the test shows smaller, but still present impact of municipalities adjacent to the capital. The number of clusters proposed by this methodology (five) is shown to be acceptable, but not optimal.

Table 1. **Statistical parameters of medoids representing their respective clusters in accordance to the chosen indicator sets (2011-2020 data)**

Indicator set based on	Cluster (see descriptions)	Medoid municipality	Population density (per km ²)	Share of population above working age in 2020	Population change due to natural movement in 2011-2019 (%)	Population change due to migration in 2011-2019 (%)	Population in 2020
Recano, 2017	1	Ropazu novads	61.244	0.173	0.024	0.082	32795
	2	Talsu novads	14.190	0.240	-0.044	-0.080	39006
	3	Preilu novads	12.851	0.249	-0.075	-0.090	18163
Indicator set based on	Cluster (see descriptions)	Medoid municipality	Demographic load in 2020	Share of working age population in 2020	Share of population below working age in 2020	Share of population above working age in 2020	Total population change in 2011-2019
Kuznetsova, 2015	1	Marupes novads	592.498	0.628	0.258	0.114	0.352
	2	Salaspils novads	618.284	0.618	0.181	0.202	0.029
	3	Balvu novads	596.882	0.626	0.125	0.249	-0.126
	4	Talsu novads	627.352	0.614	0.145	0.240	-0.125
	5	Dienvidkurzemes nov.	657.454	0.603	0.144	0.253	-0.145

Source: authors' elaboration

Study results show that with carefully selected data sets, a non-hierarchical cluster analysis employing k-medoids or similar method can become a feasible tool for quick and efficient classification of territorial units according to their socio-economic and demographic characteristics. Chosen clustering technique provided most useful results when carried out in several iterations, using different data cross-sections and methodologies. This includes using population dynamics over a sufficient period of time, as well as most recent static indicators showing the current state of affairs.

Both tested approaches to the categorisation of municipalities showed solid and actionable results, but did not provide a complete picture of the demographic disparities between the newly formed municipalities of Latvia. Therefore, a new, more country-specific set of clustering criteria may be developed for Latvia in the future.

Acknowledgements

This study was supported by National Research Programme “Letonica for the development of Latvian and European society” Project No. VPP Letonika-2021/4-0002 “New solutions in the study of demographic and migration processes for the development of the Latvian and European knowledge society”.

(DemoMigPro)References (*extended abstract only*)

- Dahs, A., Berzins, A. and Krumins, J., (2021). Challenges of Depopulation in Latvia's Rural Areas. In *Economic Science for Rural Development Conference Proceedings*, Issue 55. In Press.
- Hasek, O. (2020). Regionální diferenciace plodnosti podle typologie venkova. (The Regional Differentiation of Fertility by Rural Typology in Czechia). *Demografie*. Volume 62, pp. 3-13.
- Kaufman, L. and Rousseeuw, P.J., (2009). Finding groups in data: an introduction to cluster analysis. John Wiley & Sons.
- Krumins, J., Berzins, A., Dahs, A. (2020). Regional Demographic Trends in Accordance with the New Model of Territorial Division in Latvia and Future Dynamics in Statistical Regions of the Baltic States. *Economic Science for Rural Development Conference Proceedings*, Issue 54, pp. 233-240.
- Kuznetsova, T.Y., (2015). Geo-demographic typology of municipalities of the Kaliningrad region. *Mediterranean Journal of Social Sciences*, Issue 6, Vol. 6. pp. 345-350.
- Recano, J., (2017). La sostenibilidad demográfica de la España vacía. *Perspectives demográfiques*. Issue 7, pp.1-4.
- Regional Development Indicator Module, (2021). RDIM database. Available at: <http://raim.gov.lv> (accessed 16 September 2021).
- Stonawska, K., Vaishar, A. (2018). Differentiation and Typology of the Moravian Countryside. *European Countryside*. Volume 10, pp.127-140.