APPLICATION OF CLUSTER ANALYSIS ON THE DEMOGRAPHIC DEVELOPMENT OF MUNICIPALITIES IN THE DISTRICTS OF LIBERECKY REGION

Ondřej Šimpach

Abstract

The aim of the study is to examine the similarity of municipalities located in districts of Liberecky region according to various indicators from the area of demographic indicators using hierarchical cluster analysis method. Liberecky region has 4 districts in total (Ceska Lipa, Jablonec nad Nisou, Liberec and Semily), which together contain 215 municipalities. The municipalities of each district are merged into the clusters according to the selected demographic indicators using Euclidean distance coefficient. Selected attributes include the number of inhabitants, percentage of population aged 65 and over in the total population, live births, deaths total, number of immigrants and number of emigrants (all in 31 December). The hierarchical clusterization of municipalities is calculated for each district based on data of 2006 and 2011. Consequently, a comparison of the development of a five-year time interval is performed. Ascertained outputs can be used to plan community development and for urban planning such as transport and communications constructions, building of nurseries and basic schools, and for decisions about placement of cultural facilities.

Key words: Hierarchical Cluster Analysis, Euclidean distance, Demographic Indicators

JEL Code: C38, R11

Introduction

For the purposes of urban planning and decision-making on public sector investments (see Nutt, 2006), cohesion policy (see e.g. Pechrová, Kolářová, 2012) and for simplification of administrative process is advantageous, when based on certain socio-economic factors we know, how similar are certain territorial units to each other. The presented study will follow authors Lv et al. (2011), who used similar socio-economic indicators for the creation of clusters of the selected population, but their study focused more on the urban population of adults. Authors Ozus et al. (2012) used the hierarchical cluster analysis for the development of multicenter and travel patterns. They used data on population development, employed and

unemployed persons from 1970–2000 and travel statistics. The aim of this paper is to examine the similarity of municipalities located in districts of Liberecky region according to various indicators from the area of demographic indicators using hierarchical cluster analysis (see e.g. Ward, 1963), which can be used to improve the administrative decision making process (Feldstein, 1964), because there will be a precedent on how similar are the municipalities and thus can be decided about their strategy in the same way. Liberecky region has 4 districts in total (Ceska Lipa, Jablonec nad Nisou, Liberec and Semily), which together contain 215 municipalities. The hierarchical clusterization of municipalities (see e.g. Řezanková et al., 2011, Řezanková, Löster, 2013 or Arnio and Baumer, 2012) will be calculated for each district based on data of 2006 and 2011, acquired by the Ministry of Agriculture. Consequently, a comparison of the development of a 5-year time interval will be performed.

1   Methodology and Data

The distance between municipalities of each district of Liberecky region is calculated using Euclidean distance coefficient. Then the villages will be merged into 3, 4 or 5 clusters (depends on the situation), according to the selected demographic indicators using hierarchical cluster analysis and Ward's method (see e.g. Bavaud, 2010, or Danielsson, 1980). Given that the indicators that will be selected have different character and take values of different orders, it will be advantageous to normalize them using the *z*-transformation. We use the *z*-scores (see e.g. Larsen, Marx, 2000)

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

where μ is the mean of the population and σ is the standard deviation of the population. From mentioned data acquired by the Ministry of Agriculture of the Czech Republic, there were selected indicators

- *number of inhabitants,*
- *percentage of population aged 65 and over in the total population,*
- *live births,*
- *deaths total,*
- *number of immigrants* and
- *number of emigrants,*

all actual in December 31 in 2006 and 2011. The calculations and dendrograms were performed using by IBM SPSS Statistics.

2   Results

Using the *z*-transformation, hierarchical clustering based on Ward's method (Ward, 1963) and Euclidean distances, there were calculated the clusters for the district of Ceska Lipa in 2006 and 2011. Created dendrograms are shown in Figure 1, where the left chart presents the situation in 2006 and the right chart presents the situation in 2011. Clusters were determined after cutting both of two dendrograms at a distance of 8 units. This created three clusters and summarizing of these clusters for all municipalities is shown in Table 1. The left column of numbers indicates the cluster assigned in 2006, right column with numbers indicates the cluster assigned in 2011. Municipalities, which are displayed in bold, did not change their classification into clusters in five-year period. The town Ceska Lipa was not assigned to the clusters because it is the capital of the district and therefore outlier.

Fig. 1: Hierarchical Clusters - Dendrogram using Ward's linkage for district Ceska Lipa in 2006 (left) and 2011 (right), Euclidean distances.
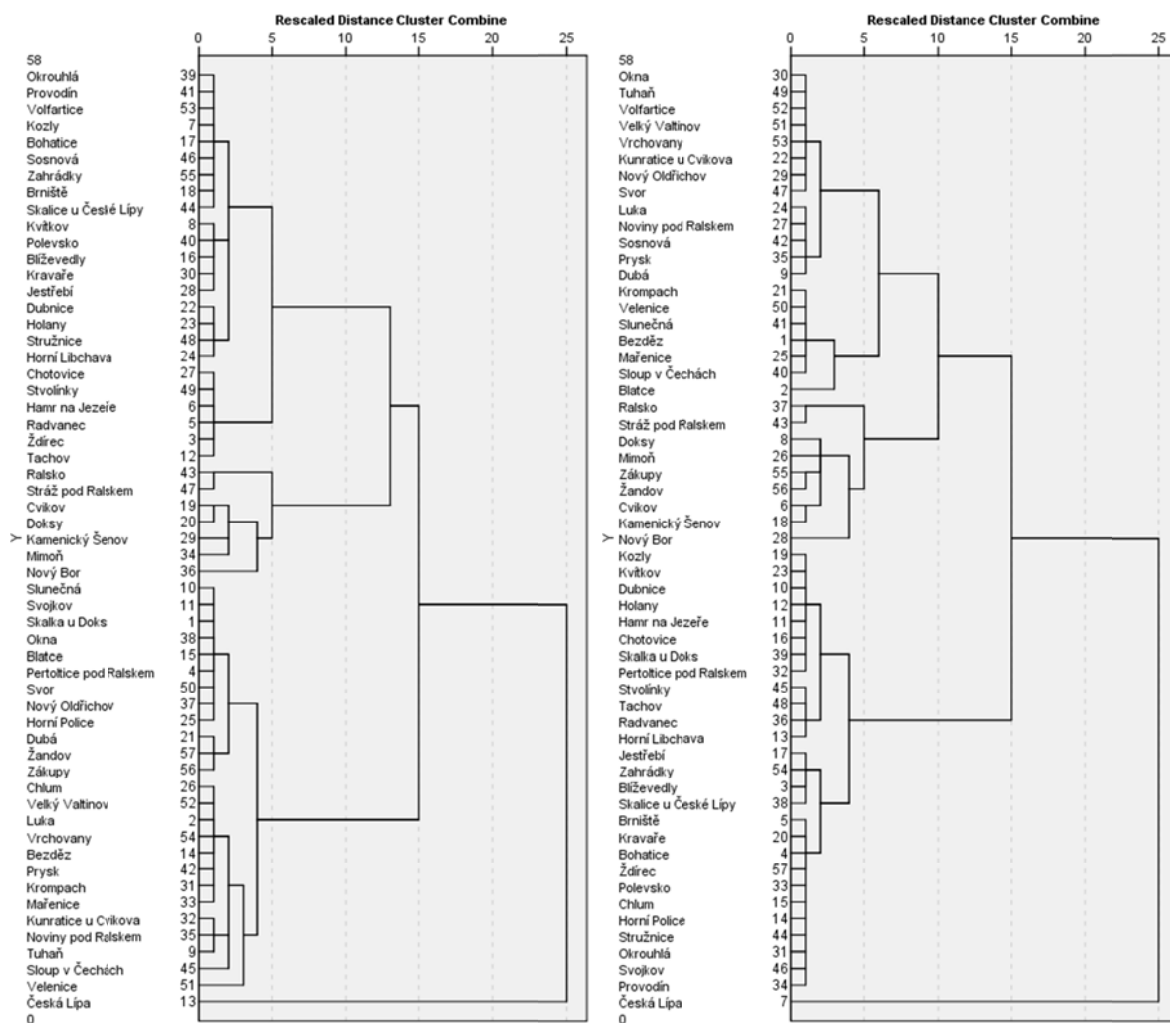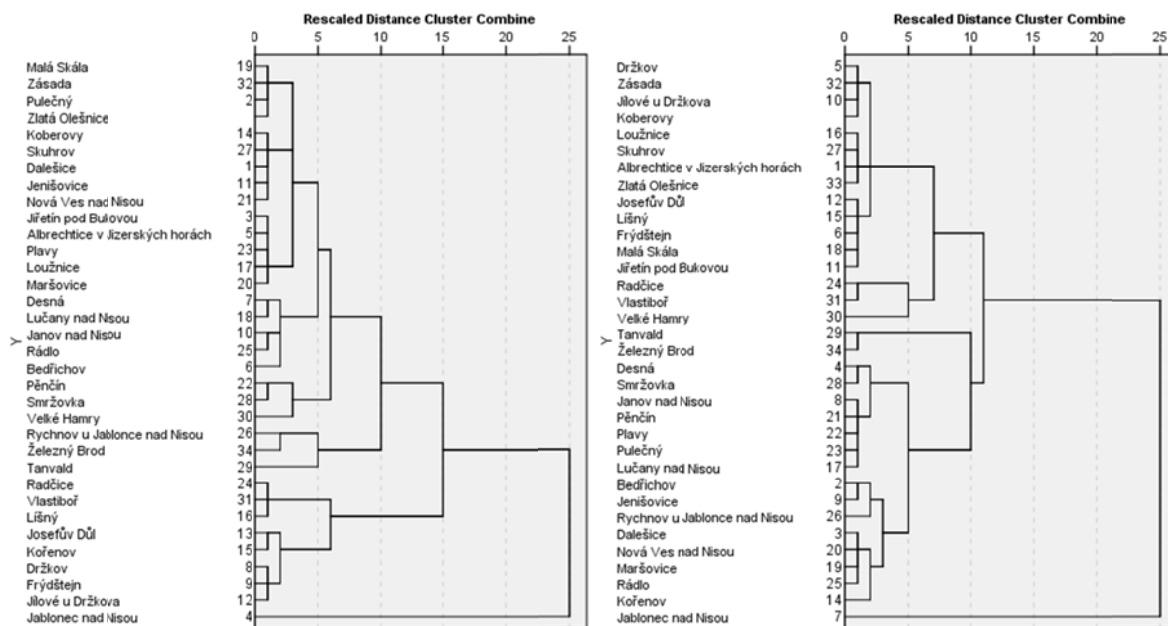

Source: author's calculations

Tab. 1: The numbers of clusters assigned to municipalities in the district Ceska Lipa in 2006 (left column) and 2011 (right column)

Blíževedly	1	3	Stvolínky	1	3	Luka	3	1
Bohatice	1	3	Tachov	1	3	Mařenice	3	1
Brniště	1	3	**Volfartice**	1	1	Noviny pod Ralskem	3	1
Dubnice	1	3	Zahrádky	1	3	Nový Oldřichov	3	1
Hamr na Jezeře	1	3	Ždírec	1	3	Okna	3	1
Holany	1	3	**Cvikov**	2	2	**Pertoltice pod Ralskem**	3	3
Horní Libchava	1	3	**Doksy**	2	2	Prysk	3	1
Chotovice	1	3	**Kamenický Šenov**	2	2	**Skalka u Doks**	3	3
Jestřebí	1	3	**Mimoň**	2	2	Sloup v Čechách	3	1
Kozly	1	3	**Nový Bor**	2	2	Slunečná	3	1
Kravaře	1	3	**Ralsko**	2	2	**Svojkov**	3	3
Kvítkov	1	3	**Stráž pod Ralskem**	2	2	Svor	3	1
Okrouhlá	1	3	Bezděz	3	1	Tuhaň	3	1
Polevsko	1	3	Blatce	3	1	Velenice	3	1
Provodín	1	3	Dubá	3	1	Velký Valtinov	3	1
Radvanec	1	3	**Horní Police**	3	3	Vrchovany	3	1
Skalice u České Lípy	1	3	**Chlum**	3	3	Zákupy	3	2
Sosnová	1	1	Krompach	3	1	Žandov	3	2
Stružnice	1	3	Kunratice u Cvikova	3	1			

Source: author's calculations

The dendrograms for the district of Jablonec nad Nisou in 2006 and 2011 are shown in Figure 2, where the left chart again presents the situation in 2006 and the right chart in 2011.

Fig. 2: Hierarchical Clusters - Dendrogram using Ward's linkage for district Jablonec nad Nisou in 2006 (left) and 2011 (right), Euclidean distances.
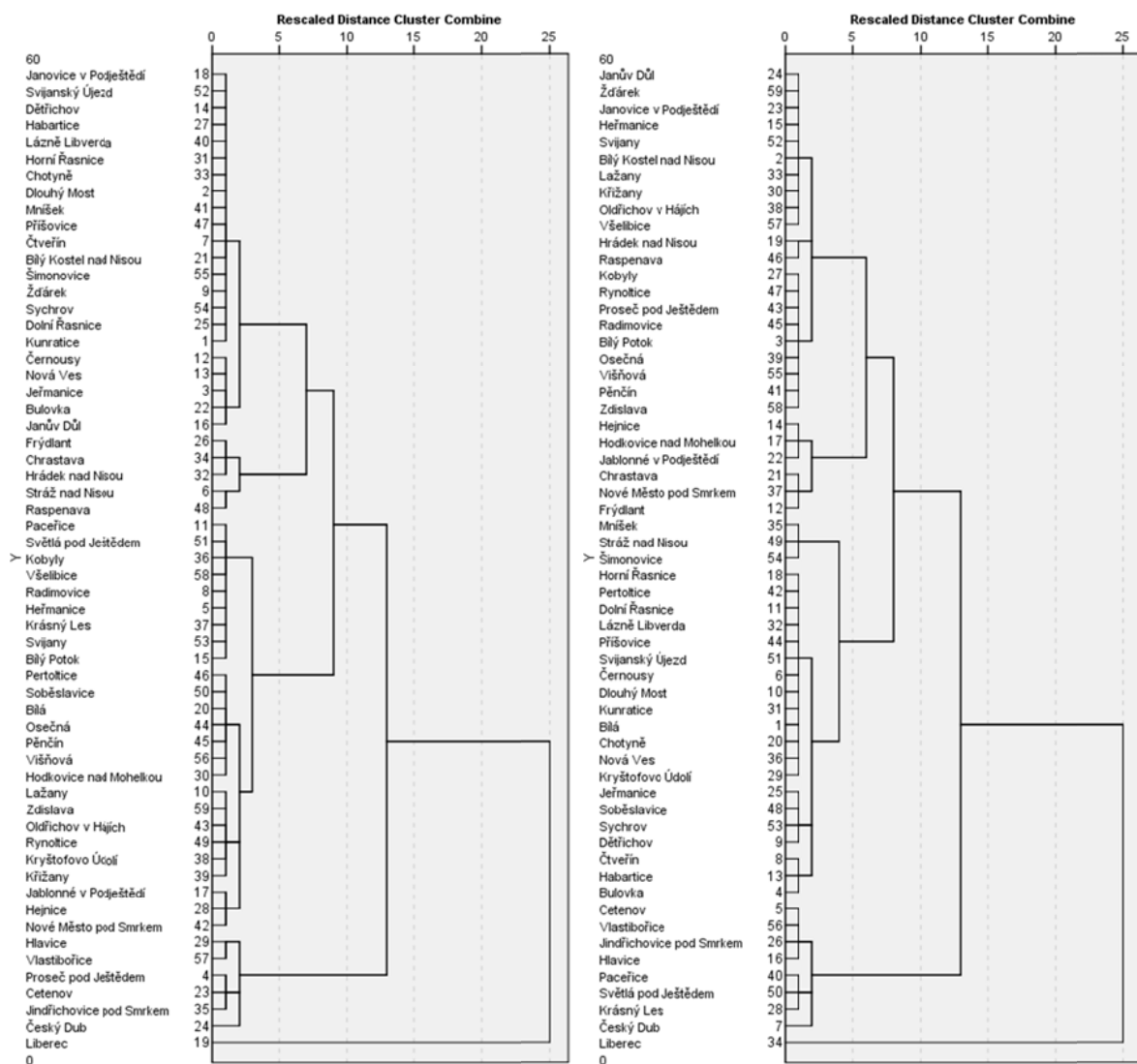

Source: author's calculations

Tab. 2: The numbers of clusters assigned to municipalities in the district Jablonec nad Nisou in 2006 (left column) and 2011 (right column)

Malá Skála	1	1	Maršovice	1	3	Rychnov u Jablonce nad Nisou	2	3
Albrechtice v Jizerských horách	1	1	Nová Ves nad Nisou	1	3	**Tanvald**	2	2
Bedřichov	1	3	Pěnčín	1	3	**Železný Brod**	2	2
Dalešice	1	3	Plavy	1	3	Držkov	3	1
Desná	1	3	Pulečný	1	3	Frýdštejn	3	1
Janov nad Nisou	1	3	Rádlo	1	3	Jílové u Držkova	3	1
Jenišovice	1	3	**Skuhrov**	1	1	Josefův Důl	3	1
Jiřetín pod Bukovou	1	1	Smržovka	1	3	**Kořenov**	3	3
Koberovy	1	1	**Velké Hamry**	1	1	Líšný	3	1
Loužnice	1	1	**Zásada**	1	1	Radčice	3	1
Lučany nad Nisou	1	3	**Zlatá Olešnice**	1	1	Vlastiboř	3	1

Source: author's calculations

Fig. 3: Hierarchical Clusters - Dendrogram using Ward's linkage for district Liberec in 2006 (left) and 2011 (right), Euclidean distances.

Source: author's calculations

After cutting both dendrograms in Figure 2 at a distance of 8 units, we can observe 3 clusters in total. The summarizing of these 3 clusters for all municipalities in district Jablonec nad Nisou is shown in Table 2 and the municipalities, which are displayed in bold, did not change their classification into clusters in five-year period. It is interesting to note that of 56 municipalities of the district Ceska Lipa 14 did not change their classification into clusters (which is relatively 25 %), and of 33 municipalities of the district Jablonec nad Nisou 12 did not change their classification (which is relatively 36.36 %). The town Jablonec nad Nisou was not considered to the clusters because it is the capital of the district, therefore outlier as in the case of Ceska Lipa.

The second most numerous district in the Liberecky region in terms of numbers of municipalities is district Liberec. Dendrograms for the year 2006 and 2011 are shown in Figure 3. Due to the higher number of municipalities in the district, the dendrograms were cut at a distance of 7 units. This led to the creation of 4 clusters, which are clearly shown in Table 3. Municipalities, which did not change their classification into clusters in five-year period, are displayed in bold. There are 15 out of 60, which is relatively 25 % (the same relatively proportion as in district Ceska Lipa). The town Liberec was not considered as in the case of Ceska Lipa and Jablonec nad Nisou, because it is the capital of the district, therefore outlier.
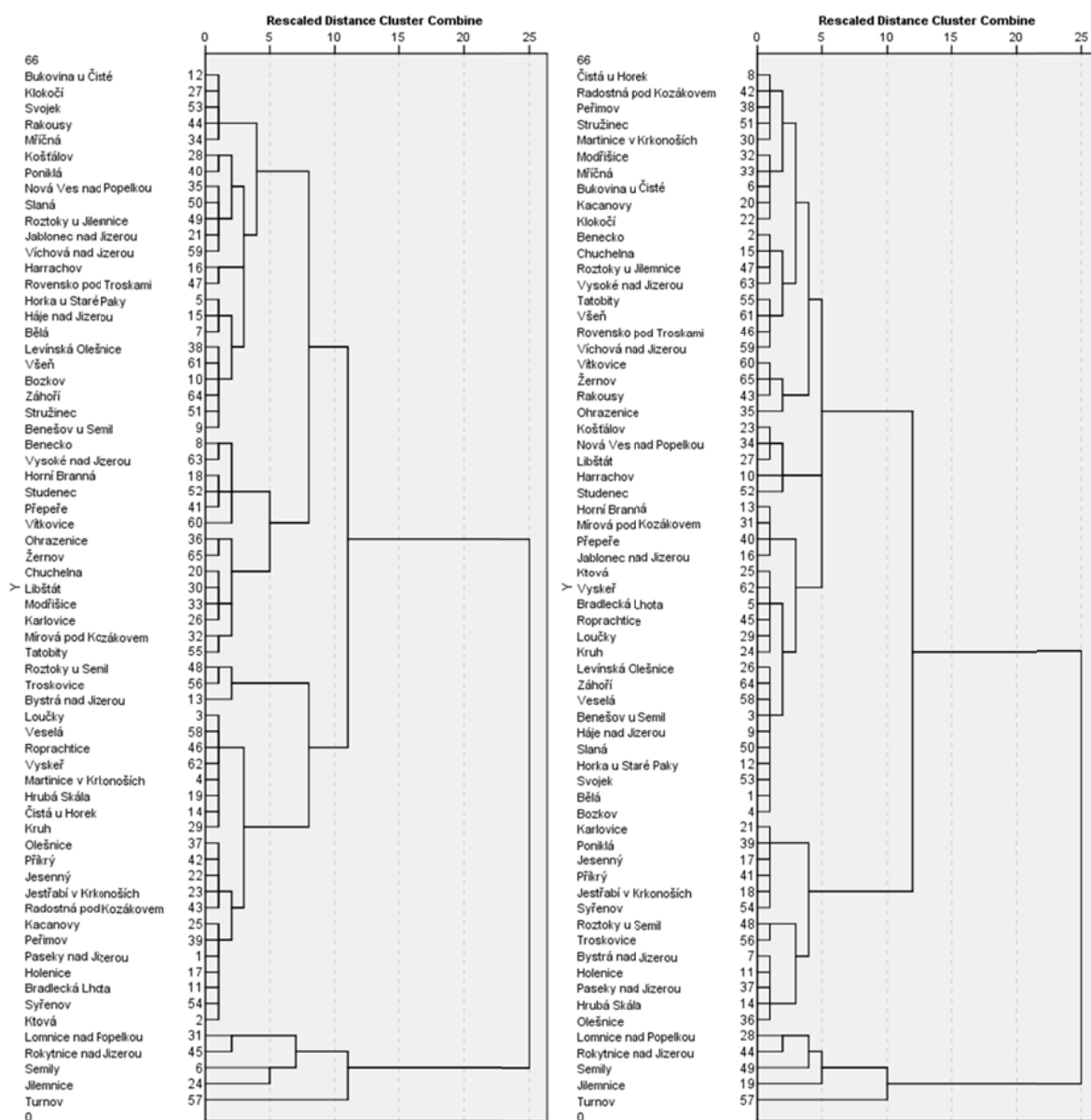
Tab. 3: The numbers of clusters assigned to municipalities in the district Liberec in 2006 (left column) and 2011 (right column)

Bílý Kostel nad Nisou	1	1	Šimonovice	1	3	Oldřichov v Hájích	3	1
Bulovka	1	3	**Žďárek**	1	1	Osečná	3	1
Černousy	1	3	**Frýdlant**	2	2	Pěnčín	3	1
Čtveřín	1	3	Hrádek nad Nisou	2	1	**Pertoltice**	3	3
Dětřichov	1	3	**Chrastava**	2	2	Radimovice	3	1
Dlouhý Most	1	3	Raspenava	2	1	Rynoltice	3	1
Dolní Řasnice	1	3	Stráž nad Nisou	2	3	**Soběslavice**	3	3
Habartice	1	3	Pacéřice	3	4	Světlá pod Ještědem	3	4
Horní Řasnice	1	3	**Bílá**	3	3	Svijany	3	1
Chotyně	1	3	Bílý Potok	3	1	Višňová	3	1
Janovice v Podještědí	1	1	Hejnice	3	2	Všelibice	3	1
Janův Důl	1	1	Heřmanice	3	1	Zdislava	3	1
Jeřmanice	1	3	Hodkovice nad Mohelkou	3	2	**Cetenov**	4	4
Kunratice	1	3	Jablonné v Podještědí	3	2	**Český Dub**	4	4
Lázně Libverda	1	3	Kobyly	3	1	**Hlavice**	4	4
Mníšek	1	3	Krásný Les	3	4	**Jindřichovice pod Smrkem**	4	4
Nová Ves	1	3	**Kryštofovo Údolí**	3	3	Proseč pod Ještědem	4	1
Příšovice	1	3	Křižany	3	1	**Vlastibořice**	4	4
Svijanský Újezd	1	3	Lažany	3	1			
Sychrov	1	3	Nové Město pod Smrkem	3	2			

Source: author's calculations

The district Semily is in terms of numbers of municipalities the most numerous of the considered districts. It has 66 municipalities, which were divided into the clusters in 2006 and 2011. While in 2006 the homogeneity was lower, in 2011 were the municipalities much more similar. Dendrograms for the year 2006 (left chart) and 2011 (right chart) are shown in Figure 4. After cutting both of two dendrograms at a distance of 8 units, we can observe 5 clusters in total in 2006 and only 3 clusters in total in 2011. Therefore, the changes in numbers of clusters are not relevant and the differences in numbers of clusters (Table 4) are not significant.

Fig. 4: Hierarchical Clusters - Dendrogram using Ward's linkage for district Semily in 2006 (left) and 2011 (right), Euclidean distances.

Source: author's calculations

Tab. 4: The numbers of clusters assigned to municipalities in the district Semily in 2006 (left column) and 2011 (right column)[1]

Municipality	2006	2011	Municipality	2006	2011	Municipality	2006	2011
Bělá	1	1	Záhoří	1	1	Jesenný	4	2
Benešov u Semil	1	1	Benecko	2	1	Jestřabí v Krkonoších	4	2
Bozkov	1	1	Horní Branná	2	1	Kacanovy	4	1
Bukovina u Čisté	1	1	Chuchelna	2	1	Kruh	4	1
Háje nad Jizerou	1	1	Karlovice	2	2	Ktová	4	1
Harrachov	1	1	Libštát	2	1	Loučky	4	1
Horka u Staré Paky	1	1	Mírová pod Kozákovem	2	1	Martinice v Krkonoších	4	1
Jablonec nad Jizerou	1	1	Modřišice	2	1	Olešnice	4	2
Klokočí	1	1	Ohrazenice	2	1	Paseky nad Jizerou	4	2
Košťálov	1	1	Přepeře	2	1	Peřimov	4	1
Levínská Olešnice	1	1	Studenec	2	1	Příkrý	4	2
Mříčná	1	1	Tatobity	2	1	Radostná pod Kozákovem	4	1
Nová Ves nad Popelkou	1	1	Vítkovice	2	1	Roprachtice	4	1
Poniklá	1	2	Vysoké nad Jizerou	2	1	Syřenov	4	2
Rakousy	1	1	Žernov	2	1	Veselá	4	1
Rovensko pod Troskami	1	1	Bystrá nad Jizerou	3	2	Vyskeř	4	1
Roztoky u Jilemnice	1	1	Roztoky u Semil	3	2	Jilemnice	5	3
Slaná	1	1	Troskovice	3	2	Lomnice nad Popelkou	5	3
Stružinec	1	1	Bradlecká Lhota	4	1	Rokytnice nad Jizerou	5	3
Svojek	1	1	Čistá u Horek	4	1	Semily	5	3
Víchová nad Jizerou	1	1	Holenice	4	2			
Všeň	1	1	Hrubá Skála	4	2			

Source: author's calculations

In the case of district Semily it is interesting, that the town Semily itself, although being the capital town of the district, is not an outlier. Outlier in this case is town Turnov, which did not fall in any cluster. The increment in the homogeneity of the considered municipalities in the case of district Semily will help with the local planning and administrative decisions-making in the future, because the municipalities will be able to be similarly managed and controlled. In the past five years, the municipalities in the district Semily became similar the most of the considered districts.

Discussion and Conclusion

This study and its results are to a certain extents influenced by the chosen methodology. When using other than Ward's method for clustering, we would get completely different results and clusters (see e.g. Löster, 2012). For example single linkage (i.e. nearest neighbour) would create less balanced clusters in terms of the number of municipalities included.

---

[1] Note that the changes in numbers of clusters are not relevant in the case of district Semily. In 2006, the municipalities were divided into five clusters and in 2011 into three clusters, because more municipalities were similar in observed characteristics. Therefore the differences in numbers of clusters are not significant.

Similarly we could have obtained different results if the input data were not normalized (but the non-normalization of data was not correct approach in this case). The more the individual municipalities in the region are similar in the terms of demographic indicators, the easier is subsequently the official planning, because it allows to access to individual municipalities in a similar approach (see more in Bartošová, Želinský, 2013). In this study the most significant homogenization of municipalities occurred in the case of district Semily.

This paper could be extendet in future by other indicators used for clusterization of the municipalities e.g. by unemployment (see e.g. Löster and Langhamrová, 2011) and other demographic indicators (e.g. fertility rate, see Fiala and Langhamrová, 2012), which are also related to the reproduction of the population.

Acknowledgment

References

Arnio, Ashley N., Baumer, Eric P. (2012): Demography, foreclosure, and crime: Assessing spatial heterogeneity in contemporary models of neighborhood crime rates, *Demographic Research*, Vol. 26, (May 2012), p. 449-486.

Bartošová, J., Želinský, T. (2013): Extent of poverty in the Czech and Slovak Republics fifteen years after split. *Post-Communist Economies*, Vol. 25, Iss. 1, 2013, p. 119 – 131.

Bavaud, F. (2010): Euclidean Distances, Soft and Spectral Clustering on Weighted Graphs, *Machine Learning and Knowledge Discovery in Databases Lecture Notes in Computer Science*, Volume 6321, 2010, pp 103-118.

Danielsson, Per-Erik (1980): Euclidean distance mapping, *Computer Graphics and Image Processing*, Volume 14, Issue 3, November 1980, Pages 227–248.

Feldstein, Martin S. (1964): Net Social Benefit Calculation and the Public Investment Decision, *Oxford Economic Papers New Series*, 16 (1) (Mar., 1964), pp. 114-131.

Fiala, T., Langhamrová, J. (2012): What Rate of Fertility and Extent of Migration Would Be Needed for Stable Population Development in the Czech Republic in This Century? *Demografie*, 2012, roč. 54, č. 4, s. 382–404. ISSN 0011-8265.

Larsen, Richard J. and Marx, Morris L., (2000): *An Introduction to Mathematical Statistics and Its Applications*, Third Edition, p. 282, ISBN 0-13-922303-7.

Löster, T. (2012): Kritéria pro hodnocení výsledků shlukování se známým zařazením do skupin založená na konfuzní matici. *Forum Statisticum Slovacum [online]*, 2012, roč. 8, č. 7, s. 85–89. ISSN 1336-7420. URL: http://ssds.sk/casopis/archiv/2012/fss0712.pdf.

Löster, T., Langhamrová, J., (2011): *Analysis of long-term unemployment in the czech republic* . In Loster Tomas, Pavelka Tomas (Eds.), *International Days of Statistics and Economics* (pp. 307-316). ISBN 978-80-86175-77-5.

Lv, J., Liu, QM., Ren, YJ., Gong, T., Wang, SF. and Li, LM., (2011): Socio-demographic association of multiple modifiable lifestyle risk factors and their clustering in a representative urban population of adults: a cross-sectional study in Hangzhou, China, *International Journal of Behavioral Nutrition and Physical Activity*, 2011, Vol. 8 : 40.

Nutt, Paul, C. (2006): Comparing Public and Private Sector Decision-Making Practices, *Journal of Public Administration Research and Theory*, (April 2006) 16 (2): 289-318.

Ozus, E., Akın, D., and Çiftçi, M. (2012): Hierarchical Cluster Analysis of Multicenter Development and Travel Patterns in Istanbul, *Journal of Urban Planning and Development*, Volume **138**, Issue 4 (December 2012), 303–318.

Pechrová, M., Kolářová, A., (2012): Does the cohesion policy mitigate the disparities among the regions in the Czech Republic?. Karviná 09.11.2012. *In: Mezinárodní vědecká konference doktorandů a mladých vědeckých pracovníků [CD-ROM]*. Opava : Slezská univerzita, 2012, p. 224–234. ISBN 978-80-7248-800-1.

Řezanková, H., Löster, T., Húsek, D., (2011): Evaluation of Categorical Data Clustering. Fribourg 26.01.2011–28.01.2011. *In: Advances in Intelligent Web Mastering – 3*. Berlin : Springer Verlag, 2011, s. 173–182. ISBN 978-3-642-18028-6. ISSN 1867-5662.

Řezanková, H., Löster, T., (2013): Shluková analýza domácností charakterizovaných kategoriálními ukazateli. *E+M. Ekonomie a Management*, 16 (3), 139-147. ISSN: 1212-3609

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

Contact

Ondřej Šimpach

University of Economics in Prague, Faculty of Informatics and Statistics

Winston Churchill sq. 4, 130 67 Prague 3 – Žižkov, Czech Republic

ondrej.simpach@vse.cz