Assessment of Participation in Cultural Activities in Poland by Selected Multivariate Methods

Alicja Grześkowiak

Wrocław University of Economics

alicja.grzeskowiak@ue.wroc.pl

Abstract

This paper presents analyses of participation in cultural activities by Poles. The analyses are carried out on the base of metric and non-metric data retrieved from the Eurobarometer survey. The study includes two main aspects: the comparison of the involvement in Poland with the situation in other European Union countries and the detection of similarities among various form of the engagement. The socio-economic background of the respondents is also taken into account, namely age, gender, place of residence and level in the society. Chosen multivariate methods are applied to identify regularities in the participation. The results of the analyses are presented graphically to facilitate the interpretation. A two-step procedure is used for a better understanding of the participation schemes. The first step includes the partition of qualitative variables into relatively homogenous groups leading to the reduction of the multidimensionality. The second step is focused on evaluating the participation with respect to the variables forming the identified clusters.

Keywords: participation in cultural activities, clustering methods, categorical data.

1. Introduction

The access to culture plays an important role across Europe but still an essential part of the population is not widely involved in cultural activities despite many Council conclusions on its importance, namely in terms of combating poverty and social exclusion or developing creative and intercultural competences (European Union 2012, p. 5). A general conclusion of a rather low level of the participation in Poland can be drawn from the Eurobarometer Reports (European Commission 2007, 2013). A vast national survey concerning this area was carried out in 2009 (GUS 2012). Other comprehensive national studies under a common name *Social Diagnosis* deal with this issue mostly in terms of the cultural needs and financial obstacles to the involvement (Czapiński & Panek 2015).

The main objective of this study is to identify and analyze patterns of the participation in cultural activities by Poles according to the Eurobarometer most recent data (2013). Some specific objectives are also formulated:

- to compare the engagement in Poland with the situation in other UE countries,
- to detect the similarities in the participation in various forms and to compare them with chosen socio-economic characteristics,
- to apply multivariate methods adequate for the variables of various type (metric, non-metric) in order to group either units or variables,
- to support the interpretation of the results of the analyses by visualization techniques.

Selected clustering methods are applied to identify regularities in the participation in cultural activities. The results of the analysis are presented graphically by dendrograms, barcharts and heatmap plots. The analyses are carried out on the base of *Eurobarometer* survey outcomes allowing for making international comparisons due to the unified process of data collection.

2. Data description

This study is based on data from Special Eurobarometer survey[1] devoted to cultural participation of European Union citizens. The survey was requested by the European Commission and carried out in April and May 2013. The questionnaire included a question concerning the frequency of the participation in various cultural activities in the last year. The literal question was as follows: "*How many times in the last twelve months have you…?*" (TNS Opinion 2013). Nine options were given for the evaluation, i.e. "*seen a ballet, a dance performance or an opera; been to the cinema; been to the theatre; been to a concert; visited a public library, visited a historical monument or site (palaces, castles, churches, gardens, etc.); visited a museum or gallery, watched or listened to a cultural programme on TV or on the radio, read a book*" (TNS Opinion 2013). As the original descriptions of the activities are rather long, shorter versions of them are used in further considerations in order to make the visualizations and tables more clear: Opera, Cinema, Theatre, Concert, Library, Monument, Museum, RTV, Book, respectively. The respondents of the survey were asked to choose one from possible answers: "*not in the last 12 months; 1-2 times; 3-5 times; more than 5 times*". The analyses of the participation are based both on aggregated and individual data. The international comparison was carried out with respect to the percentages of respondents who declared that they had took part in a particular activity at least once in the last twelve months. The analysis of the behaviour of the Polish respondents was performed on individual, categorical data. The answers to the question about the involvement in various activities were binarized (0 - no participation at all, 1 - participation at least once in the last year). A set of socio-economic non-metric variables was also taken into consideration:

- gender: male, female;
- place of residence: rural area or village, small/medium-sized town, large town/city;
- age (categorized): 15-24, 25-39, 40-54, 55+;
- level in the society (self placement on ten-degree scale, categorized): low (1-4), middle (5-6), high (7-10).

As some missing values were detected, some observations had to be omitted and the final dataset comprised N = 960 cases.

3. Analytical methods

According to the objectives of this research, some selected multivariate techniques are applied to disclose relationships and patterns in datasets. Various clustering procedures as well as certain visualization methods supporting the interpretation of the results are used. The clustering algorithms were chosen as they constitute crucial methods of the scientific inquiry, especially in social sciences when no particular underlying theory of the phenomenon is available and the goal is to search for and to reveal the existing patterns (Bartholomew et al. 2008, p.18).The principal objective of cluster analysis is to assign individuals (observations, units) to clusters when the group membership is not known *a priori* (Afifi, May & Clark 2003). There are two main types of clustering algorithms: partitioning and hierarchical (Rencher 2003, p.452). Only the latter approach is used in the analyses, so its general idea is briefly presented. Hierarchical clustering is done in a few predefined steps, namely: (1) collecting a data matrix representing the objects and the attributes describing them, (2) standardizing the data matrix if necessary, (3) measuring the similarities among all pairs of objects, (4) applying a specific method to find the hierarchy of the similarities among objects and to present it in form of dendrogarm (Romesburg 2004, p.3). Detailed descriptions of numerous clustering algorithms can be found in many publications, e.g. Anderberg (1973), Aggarwal & Reddy (2013), Everitt et al. (2011). Although the most common purpose of the cluster analysis is to group the units, the same procedures may be applied to group the variables according to their mutual behaviour and to reveal structures and "natural associations" among variables within complex datasets (Anderberg 1973). Moreover, some specific methods are proposed for variables clustering only. As indicated by Chavent et al. (2013) the most important methods for metric data are VARCLUS procedure implemented in SAS software, CLV method (Vigneau & Qannari 2003) and diametrical clustering (Dhillon, Marcotte & Roshan 2003). As survey data are often non-metric in nature, these techniques cannot be used in many social science studies based on such data. Another solution dealing with both quantitative and qualitative data is proposed by Chavent et al. (2013) and implemented in ClustOfVar R package. The clustering procedures are based on a principal components method appropriate for a mixture of qualitative and quantitative

variables and maximize a homogeneity criterion - the degree of the association with the central quantitative synthetic variable measured either by correlation coefficient or correlation ratio (Chavent et al. 2013). The stability of the partitions may be assessed by Rand (Rand 1971) or adjusted Rand (Hubert & Arabie 1985) criteria.

The results of the hierarchical clustering are usually presented graphically by dendrograms but another extended approach is possible. A visualization technique called *cluster heatmap* is used to show or identify the relationships between the units and the variables with respect to the clustering outcomes. This visualization method is widely used in biological research, mostly to data collected from microarrays but there are not any obstacles to apply this technique to other data (Pryke, Mostaghim & Nazemi 2007). The cluster heatmap consists of a rectangle representing the data matrix with dendrograms attached to its margins and it facilitates the examination of row, column, and joint cluster configuration (Wilkinson & Friendly 2009). The rectangle is divided into cells whose colours reflect the values of the original dataset; the columns and the rows are permuted in order to properly show the clustering of the variables and the units, respectively (Chen, Härdle & Unwin 2007, p. 567). An interesting heatmap presentation with a variety of options is implemented in pheatmap R package (Kolde, 2015).

4. Participation in cultural activities in Poland on the European background

The comparison of Poles' participation in cultural activities with the patterns observed in all European Union member states (at the moment when the survey took place) was performed on the base of the variables representing the percentages of respondents who declared that they had took part in a particular activity at least once in the last twelve months. Hence, the data matrix comprised 27 objects described by 9 variables (attributes). The input data in this case were metric so an agglomerative clustering algorithm was applied. The data were standardized. The Euclidean distance was chosen as the measure of dissimilarities among the pairs of objects and Ward's method was selected as the criterion for merging clusters in the hierarchical procedure. Finally, a heatmap presentation was used to visualize the outcomes of the analysis and to facilitate the interpretation of the patterns. The heatmap reflecting the standardized values of the analyzed data and the clustering results is given in Figure 1. The position of Poland is marked by an arrow.

The heatmap in Figure 1 shows the partition of the countries into four clusters. The first cluster (as seen from the top of the figure) consists of the worst performers in terms of the participation in cultural activities: Bulgaria, Hungary, Poland, Greece, Cyprus, Portugal and Romania. All participation indicators in these countries are below the average. A completely different pattern can be noticed in the countries who constitute the second cluster: Sweden, Denmark and the Netherlands where the engagement in cultural activities is the highest across the European Union. The third cluster comprising Czech Republic, Slovakia, Italy, Spain and Malta is characterized by the values lower than or close to the average. The fourth and the biggest cluster contains the other member states (not listed above) and can be described as moderate as the values are higher or close to the average.

Some similarities among the variables can be also indicated, particularly among visiting museums, galleries and monuments, reading books and going to the cinema. Other regularities may be noticed between being to a concert or being to a theatre as well as between seeing a ballet, a dance performance or an opera and watching or listening to cultural programme on TV or on the radio.
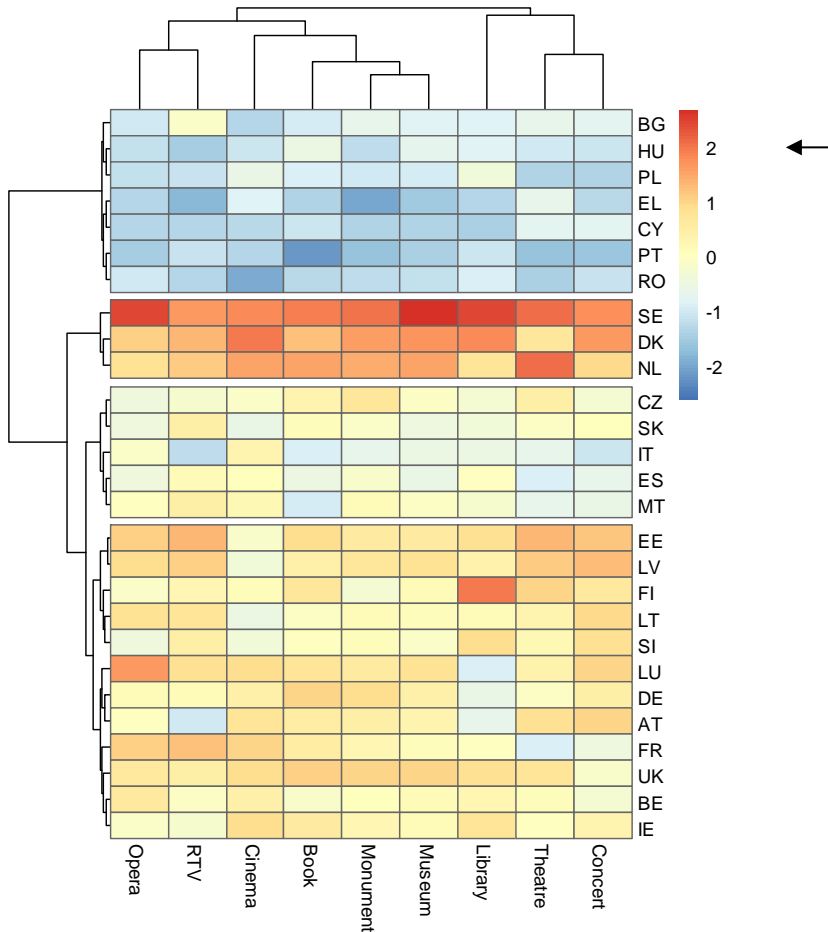
Figure 1. Results of the agglomerative hierarchical clustering (Ward's method) of the EU countries and the variables describing the participation in the cultural activities.

Note: The position of Poland is highlighted by the arrow. Abbreviations: AT- Austria, BE - Belgium, BG - Bulgaria, CY - Cyprus, CZ - Czech Republic, DE - Germany, DK - Denmark, FR - France, HU - Hungary, EE - Estonia, EL - Greece, ES - Spain, FI - Finland, IE - Ireland, IT - Italy, LT - Lithuania, LU - Luxembourg, LV - Lativa, , MT - Malta, NL - Netherlands, PL - Poland, PT - Portugal, RO - Romania, SE - Sweden, SI - Slovenia, SK - Slovakia, UK - United Kingdom.

Source: own elaboration in pheatmap R package on the base of Special Eurobarometer 79.2 (399) data.

As it can be seen from the recognized clusters and patterns, the participation in cultural activities in Poland is among the lowest in the European Union. This unfavorable situation induces the need for a more detailed analysis based on individual data and with respect to the socio-economic background.

5. Participation in cultural activities in Poland - analysis of non-metric data

The involvement in cultural activities in Poland was evaluated on the base of nine categorical variables describing various aspects of the phenomenon. In the first step of the analysis it was verified whether there are patterns due to the type of the

participation. For this purpose, a method available in the ClustOfVar R package (Chavent et al. 2013) was used that allows detecting clusters among qualitative variables. The agglomeration process is illustrated by the means of a dendrogram in Figure 2. Figure 3 shows the evaluation of the stability of the dendrogram partitioning on the base of the mean adjusted Rand criterion calculated from 100 bootstrap samples. The highest index corresponds to the division into eight clusters of variables but this solution is unfortunately not informative. Therefore, the split into three clusters was taken into consideration, for which the Rand criterion was the second largest.
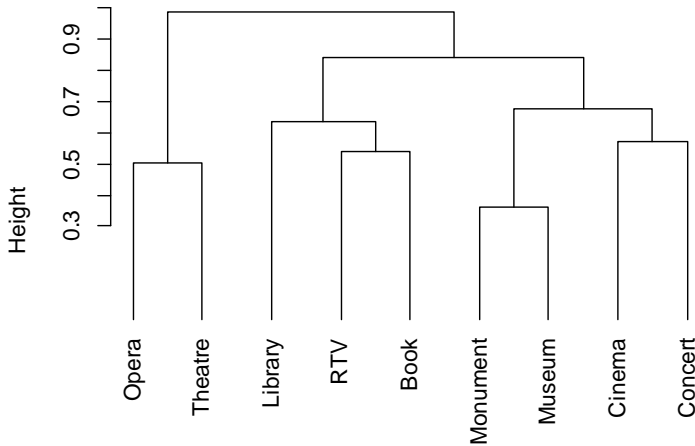

Figure 2. Dendrogram representing the clustering of the variables describing the participation in the cultural activities by Poles

Source: own elaboration in ClustOfVar R package on the base of Special Eurobarometer 79.2 (399) data.
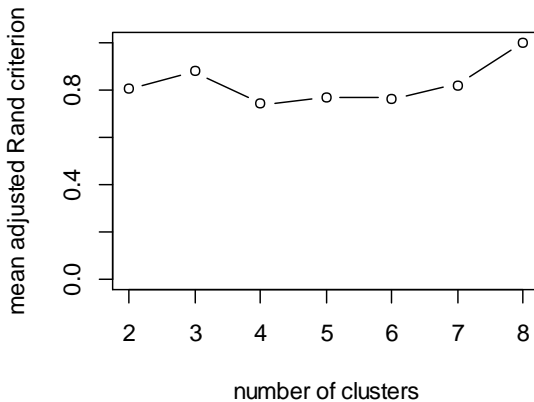

Figure 3. The evaluation of the stability of the dendrogram partitions according to the adjusted Rand criterion (calculated from 100 bootstrap samples).

Source: own elaboration in ClustOfVar R package on the base of Special Eurobarometer 79.2 (399) data.

Three identified clusters are as follows: (1) *seeing a ballet, a dance performance or an opera and being to the theatre*, (2) *being to a concert, being to the cinema, visiting a historical monument or site, visiting a museum or gallery*, (3) *visiting a public library, watching or listened to a cultural programme on TV or on the radio, reading a book.* It is worth underlying that the clusters are different in nature. The first one seems to be the most sophisticated and comprises some events available only in large towns or cities. The third one consists of easily accessible and low-cost activities as reading books, listening

to the radio, watching TV, visiting a public library. The second cluster includes the activities carried out outside home and probably requiring more time and financial expenditures. The homogeneity of the clusters can be evaluated by the degree of association between the variables constituting the cluster and the central synthetic variable; in the case of qualitative variables correlation ratio is applied for this purpose (Chavent et al. 2013). The results in Table 1 show that the partition into three clusters is reasonable as the correlation ratios are relatively high.

Table 1. Associations between the input variables and the clusters' synthetic variables

Cluster	Variable	Loading (cr)
1	Opera	0,75
	Theatre	0,75
2	Cinema	0,53
	Concert	0,52
	Monument	0,68
	Museum	0,66
3	Library	0,56
	RTV	0,57
	Book	0,69

Source: own elaboration in ClustOfVar R package on the base of Special Eurobarometer 79.2 (399) data.

The identification of three different clusters gives reason to perform the analyzes separately for each of them. It is an alternative, more detailed approach than the *Index of cultural practice* proposed in (European Commission 2013, p. 9), which has many advantages such as the simplicity of construction and interpretation, but treats all cultural activities equally. In the case of Poland the approach based on pre-clustering appears to be justified because of large differences in participation rates calculated separately for the three recognized groups. The particular terms are defined as follows: full participation (F) - participation in all activities within the cluster, partial participation (P) - participation in at least one activity within the cluster but not in all of them, no participation (N) - no participation in activities within the cluster. Some indicators calculated as F, P and N ratios are given in Table 2.

Table 2. Comparison of the participation with respect to the identified clusters

Cluster	Participation indicators		
	F/N	P/N	Total: F/N+P/N
1	0,08	0,14	0,22
2	0,19	0,84	1,03
3	0,77	1,85	2,62

Source: own elaboration on the base of Special Eurobarometer 79.2 (399) data.

Essential differences in the indicators calculated for the clusters should be emphasized. In the first cluster for one person who saw a ballet, a dance performance, opera or a play in a theatre there are about five persons who did not participate in such events at all. It is the only cluster, in which the partial participation is only a bit higher than the full involvement. In the others the partial engagement is much higher than the full one. There are more people using than not using the cultural offers included in the clusters 2 and 3. Moreover, the ratio is much more favorable in the third group (2,62 as compared to 1,03). The ratio of the partial participation and no participation is higher than one only in the case of the cluster 3 which

includes easily accessible and low-cost activities. The calculated indicators show a very large diversity of Poles' participation in cultural activities depending on their type.

The next step of the analysis is the assessment of the full, partial and no participation indicators within the identified clusters with respect to a set of socio-economic variables (Figure 4).
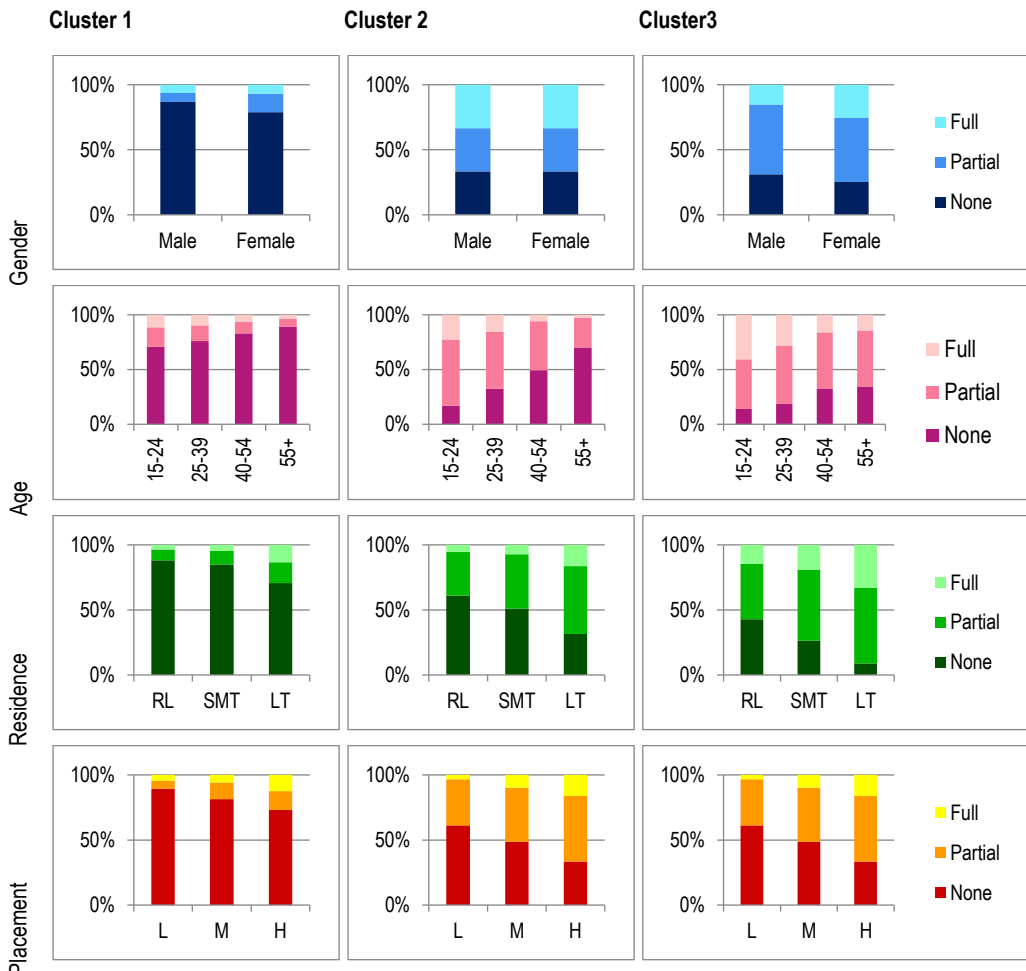

Figure 4. Participation in cultural activities with respect to the identified clusters and socio-economic characteristics.

Note: Residence: RL - Rural area or village, SMT - Small/Middle town, LT - Large town/city; Placement in the society: L - low, M - Middle, H - High.

Source: own elaboration on the base of Special Eurobarometer 79.2 (399) data.

A higher percentage of women than men declares participation in cultural activities included in the clusters 1 and 3. A vital difference can be found in the cluster 3 in the case of the full participation, which is declared by 25,6% of females and only by 15,2% of males. In contrast, the differences due to the gender are not found in the cluster 2. The age is also an important differentiating factor. Generally, the engagement in cultural activities decreases with the age of the respondents. Particularly worrying is the fact of no participation by people from the oldest age group (55+). A clear division occurs between 40+ and

younger Poles when the activities included in the cluster 3 are considered. The younger generation tends to read more books and use more the cultural offers given on the radio and TV or in the library. The place of residence also plays an important role. In the case of cluster 1 and 2, the reason may seem to be evident as the access to certain cultural events in rural areas, villages and smaller towns is limited. However, the same patterns occur in the case of the cluster 3 comprising easily accessible activities. Moreover, the gap between rural areas/villages and large towns/cities is the highest in this cluster: no participation is declared by 42,8% and 8,9% of respondents, respectively. It suggests the existence of not only objective obstacles arising from the external circumstances, but also intrinsic barriers, perhaps resulting from the lack of such needs or the lack of awareness. Some regularities combined with the perceived level in the society are also noticeable, but one must bear in mind that this assessment is a self-placement type, so it does not give objective criteria. Nevertheless, the lower placement in the society, the lower participation in cultural activities is observed. More than a half of those who place themselves at the lowest level of the society resigns from the engagement in cultural offers, irrespective of the cluster.

Conclusions

The analysis of the Eurobarometer data reveals that the participation in cultural activities in Poland is among the lowest in the European Union and varies across socio-economic factors. In particular, there are clear differences between people living in the countryside and in town/cities as well as between persons from different age groups. Very alarming, especially in the context of the aging population is no or a low involvement in cultural activities by persons 55+. It is necessary to take appropriate steps to attract older people to these forms of spending time contributing to the realization of the idea of the active aging.

The partition of variables into relatively homogenous groups in the first stage of the analysis reduces the multidimensionality of data while retaining the possibility of recognizing underlying patterns. This kind of approach is particularly helpful if groups of variables characterized by considerable similarities exist, which allows the researcher to extract meaningful clusters. Taking into account a set of the identified clusters of variables in further analysis is probably more informative than the interpretation of one composite indicator based on all variables altogether. On the other hand, the synthetic indicator has also a number of advantages - it has a simple construction, requires less complex calculations and is understood more easily by stakeholders without background in statistics. It can thus serve as an overall index for the general description of the situation, while the two-step approach giving a more detailed and more precise insight into the phenomenon can be applied in the in-depth analysis of the problem. The approach including the preliminary clustering of variables may be treated as a kind of balance between handling each factor separately, and construction of one global indicator.

Acknowledgements

References

[1] Afifi, A., May, S., Clark, V. A. (2003). *Computer-aided multivariate analysis*. CRC Press.

[2] Aggarwal, C. C., Reddy, C. K. (Eds.). (2014). *Data clustering: algorithms and applications*. CRC Press.

[3] Anderberg, M. R. (1973). *Cluster Analysis for Applications*. Academic Press, Inc., New York.

[4] Bartholomew, D. J., Steele, F., Galbraith, J., Moustaki, I. (2008). *Analysis of multivariate social science data*. CRC Press.

[5]   Chavent, M., Kuentz, V., Liquet, B., Saracco, L. (2013). Clustofvar: An R package for the clustering of variables. *Journal of Statistical Software*, 50(13), 1-16.

[6]   Chen, C. H., Härdle, W. K., Unwin, A. (2007). *Handbook of data visualization*. Springer Science & Business Media.

[7]   Czapiński J., Panek T. (Eds.). (2015). *Diagnoza społeczna 2015. Warunki i jakość życia Polaków. Rada Monitoringu Społecznego, Warszawa,* www.diagnoza.com (retrieved 2015-12-29).

[8]   Dhillon, I. S., Marcotte, E. M., & Roshan, U. (2003). Diametrical clustering for identifying anti-correlated gene clusters. *Bioinformatics*, *19*(13), 1612-1619.

[9]   European Commission. (2013). *Cultural access and participation. Report.* Special Eurobarometer 399 / Wave EB79.2 – TNS Opinion & Social, http://ec.europa.eu/public_opinion/archives/ebs/ebs_399_en.pdf (retrieved 2015-12-29).

[10] European Commission. (2007). *European Cultural Values*. Special Eurobarometer 278 / Wave 67. – TNS Opinion & Social, http://ec.europa.eu/COMMFrontOffice/PublicOpinion/index.cfm/ResultDoc/download/DocumentKy/545 72 (retrieved 2015-12-29).

[11] European Union (2012) *A report on policies and good practices in the public arts and in cultural institutions to promote better access to and wider participation in culture.* Open Method Of Coordination (OMC) Working Group Of EU Member States' Experts On Better Access To And Wider Participation In Culture http://ec.europa.eu/culture/policy/strategic-framework/documents/omc-report-access-to-culture_en.pdf (retrieved 2015-12-29).

[12] Everitt, B. S., Landau, S., Leese, M., Stahl D. (2011). *Cluster Analysis*. Wiley.

[13] GUS. (2012). *Uczestnictwo ludności w kulturze w 2009 r*. Warszawa.

[14] Hubert, L., Arabie, P. (1985). Comparing partitions. *Journal of classification*, *2*(1), 193-218.

[15] Kolde, R. (2015). *pheatmap: pretty heatmaps*. R package version 1.0.8.

[16] Pryke, A., Mostaghim, S., Nazemi, A. (2007). Heatmap visualization of population based multi objective algorithms. In *Evolutionary multi-criterion optimization* pp. 361-375. Springer Berlin Heidelberg.

[17] Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66(336), 846-850.

[18] Rencher, A. C. (2003). *Methods of multivariate analysis*. John Wiley & Sons.

[19] Romesburg, C. (2004). *Cluster analysis for researchers*. Lulu Press.

[20] TNS Opinion. (2013). *Eurobarometer 79.2. Basic bilingual questionnaire*,

[21] https://dbk.gesis.org/dbksearch/download.asp?db=E&id=52187 (retrieved 2015-12-29).

[22] Vigneau, E., Qannari, E. M. (2003). Clustering of variables around latent components. *Communications in Statistics-Simulation and Computation*, 32(4), 1131-1150.

[23] Wilkinson, L., Friendly, M. (2009). The history of the cluster heat map. *The American Statistician*, 63(2), 179-184.