Clustering Applied to the Education: A *K*-means and Hierarchical Application

Victor Daniel Gil-Vera[1] & Isabel Cristina Puerta-Lopera[2] & Catalina Quintero-Lopez[3]

[1] Teacher and Researcher, SISCO Research Group, Universidad Católica Luis Amigó, Colombia

[2] Vice-Chancellor of research, Universidad Católica Luis Amigó, Colombia

[3] Teacher and Researcher, NBA Research Group, Neurocognitive and Psychology Profiles Line, Universidad Catolica Luis Amigo, Colombia

Correspondence: Víctor Daniel Gil-Vera, Universidad Catolica Luis Amigo, Colombia. E-mail: victor.gilve@amigo.edu.co

Abstract

Currently, most schools in the world use ICT, which is why students must make use of computers and mobile devices in and out of schools. Thanks to the use of technology, students are more interested and motivated to learn, considering that motivation is one of the main engines of learning, since it encourages activity and thought. On the other hand, motivation makes students spend more time working and therefore they are more likely to learn more. The aim of this paper was to present a clustering of European countries according to the number of desktop computers available to students in primary schools (ISCED 1), lower secondary schools (ISCED 2) and upper secondary schools (ISCED 3). Was used the database developed by the ES Open Data Portal for the year 2019 on "ICT in Education". For the classification were used the hierarchical clustering and *K-means* techniques and the statistical software *Rcran 3.6.3*. These techniques were used as they have the ability to group a large number of elements into clusters, based on the similarity learned. This paper concludes that the countries with the highest GDP are not the ones that have the most desktop computers in their schools. Bulgaria is the country with the major number of desktop computers in their schools.

Keywords: ICT, education, students, learning, Europe

1. Introduction

The use of ICT in education allows the generation of new communication channels between teachers and students, encourages collaborative work, promotes reflective teaching practices, methodological updating and the acquisition of digital skills (Palomino, 2017). Thanks to the ICT use, students are permanently predisposed to interact with the computer. The versatility and interactivity offered by the ICT, the possibility of interacting virtually with other people and the large volume of scientific information available on the internet, is attracting more and more university students worldwide.

The constant participation of students encourages the development of their initiative, since they are forced to make decisions to the responses to their actions generated by the computer. In addition, autonomous, rigorous and methodical work is promoted (Talebian et al., 2014). Immediate feedback allows students to know their mistakes in real time, which facilitates the identification of weaknesses and the strategies formulation to overcome them. Thanks to ICT there is more communication between teachers and students (Arkorful & Abaidoo, 2015). The communication channels offered by the internet: e-mail, forums, blog, chat, WebQuest, etc., facilitate their interaction, since students can ask questions, share ideas, exchange resources, hold debates, etc. (Bajunid, 2001). Two of the main advantages of the ICT use in education are:

— *Cooperative learning*: the tools provided by ICT; information sources, interactive materials, e-mail, shared disk space, chat, facilitate teamwork, development of social attitudes, exchange of ideas, cooperation and personality development. Teamwork stimulates its members and makes them dialogue about the best solution for a given problem (Van Ryzin & Roseth, 2019).

— *High degree of interdisciplinarity*: the educational tasks carried out with the computer allow for a high degree of interdisciplinarity, since the computer, due to its versatility and large storage capacity, allows for various types of information processing (Lin, 2008).

Students with high IQ process information faster than those with low IQ. A high degree of innovation has been demonstrated in the activities they carry out. Due to the above, it is necessary to personalize educational intervention

strategies, in which the use of ICT is integrated, as has been demonstrated in theoretical models (Fernández-Batanero, et al., 2019). The current European education model has led to innovations in the teaching-learning process in higher education, including the incorporation of ICT. Considering the variety of learning styles that students may have, it is important that teachers personalize teaching methods and develop ICT skills (Vega-Hernández, Patino-Alonso, & Galindo-Villardón, 2018). In the European Higher Education Area (EHEA), efforts have been made to promote the use and incorporation of ICT in European Universities, emphasizing the motivational and cognitive components underlying the learning process (Valentín et al., 2013).

This paper is divided as follows: the first part presents a general contextualization of the most used clustering techniques in data analysis (hierarchical and k-means). Then the methodology used in the classification and the code used in the software R Cran 3.6.3. Finally, presents the classification results, the analysis and discussion. This paper concludes that some countries make a greater use of ICT than others and the uses differ from each other, Bulgaria is the country with the major number of desktop computers in their schools.

2. Clustering

The main task of data mining is to extract useful information from large volumes of information. Clustering is one of the most common techniques in data mining, it allows the natural discovery of groups of similar observations. In this case, the *hierarchical* and *k-means* techniques was used to group European countries according to the number of desktop computers available in the schools. This grouping can serve to identify countries that have the least number of computers in schools, so that they may be motivated to define strategies to increase these resources, considering that they are fundamental tools in education in the XXI century.

Clustering consists of an exploratory data analysis technique used to group objects that have similar characteristics. Its main use is to discover patterns that are difficult to perceive and to generate information that can contribute to decision making (Govender & Sivakumar, 2020). As mentioned above, this technique has been used to group similar data sets, observations, vectors, etc. (Jain et al., 1999). In short, it is the process of identifying groups of data (Kaufman & Rousseeuw, 2009). In fact, the objects that make up a group have a greater degree of similarity than the objects that make up other groups (Govender & Sivakumar, 2020) and also facilitates the identification of distributions and patterns of interest, making it easier to understand the underlying structure of the data (Halkidi et al., 2001). This technique was proposed in 1930, but recognition of its usefulness began in the early 1960s. There are many areas of knowledge that have used it; natural sciences, health sciences, social and human sciences, etc. (Gong & Richman, 1995).

2.1 K – Means Clustering

Unsupervised learning algorithm used to group a data set into a number (K) of groups defined by the researcher (MacQueen, 1967). Classify objects into different clusters so that objects within the same cluster are as similar as possible, while objects in different clusters are as different as possible. In the k-means clustering, each cluster is represented by a centroid that corresponds to the average number of points assigned to the cluster (Kassambara, 2017). In summary, this clustering technique associates objects in a way that minimizes variation between them (Jaiswal, 2018). Among the different types of existing average K algorithms, the most used is the one that defines the variation within the group as the sum of the squared Euclidean distances between each element of the group and the centroid (Hartigan & Wong, 1979):

$$W(Ck) = \sum x_i \in C_k \ (x_i - \mu_k)^2 \tag{1}$$

Where x_i is data belonging to the C_k cluster and μ_k is the average value of the data assigned to the C_k cluster. Each data x_i is assigned to a given cluster so that the sum of the squares of the distance of the observed data to the center of the cluster assigned μ_k is minimal (Jaiswal, 2018). Briefly, the step-by-step of this algorithm is:

— Determine the number of clusters (k) to be created.

— Randomly select (k) objects from the initial data set as the plants of the clusters.

— Assign the nearest centroid to each remaining data set, based on euclidean distance.

— For each cluster, recalculate the centroid by finding the average of all the data in the cluster.

— Iteratively minimize the total within the sum of the squares. Repeat step 3 and step 4, until the centroids do not vary or the maximum number of iterations is reached.

The total variation within the cluster or the total of the sum of squares is defined in the following equation:

$$\sum_{k=1}^{k} W(Ck) = \sum_{k=1}^{k} xi \in Ck(xi - \mu k) \tag{2}$$

This is the sum of all the clusters over the sum of the Euclidean distances squared between the data and the corresponding centroid (Kassambara, 2017).

2.2 Hierarchical Clustering

This clustering technique is more complete than the previous technique, as it generates a dendrogram (partition tree) (Xu et al., 2020). By cutting the dendrogram generated by a hierarchical grouping algorithm, different grouped results can be obtained without using it again (Xu et al., 2020). Hierarchical clustering (HC) has a temporal complexity of *O(2n)*, uses reciprocal close neighbors and reproducibility. HC starts from n clusters and successively merges the most similar ones to form a larger one. This process is repeated until the number of clusters equals the desired number. Unlike the previous technique, hierarchical clustering does not require predefined parameters. Because of this, it is the most suitable clustering technique for handling real data where setting parameters is a complex task (Xu et al., 2020; Bouguettaya et al., 2015). The key principle in this grouping technique is to repeatedly combine the two closest groups into one larger group (Pathak, 2018). The main steps of this algorithm are:

— Find the distance between each pair of points in the data set and store it in a distance matrix.

— Locate each point in its own cluster.

— Merge the closest pairs of points based on the distances from the matrix. The number of clusters is reduced by one unit.

— Re-calculate the distance between the new cluster and the old ones and store them in a new distance matrix.

— Repeat steps 2 and 3 until all the clusters are integrated into one.

Linking methods measure the distance between the clusters to decide the grouping rules. Among the different methods available are full (finds the maximum distance between clusters before merging), single (calculates the minimum distance between clusters before merging), medium (calculates the average distance between clusters before merging) and centroid (finds the centroid of cluster 1 and the centroid of cluster 2, and then calculates the distance between them before merging) (Pathak, 2018). Table 1 presents the some commonly used metrics for this clustering technique:

Table 1. Hierarchical clustering metrics

Distance	Equation
Euclidean	$\| a - b \|_2 = \sqrt{\sum_i (a_i - b_i)^2}$
Squared Euclidean	$\| a - b \|_2^2 = \sum_i (a_i - b_i)^2$
Manhattan	$\| a - b \|_1 = \sum_i | a_i - b_i|$
Maximun	$\| a - b \|_\infty = max_i \ | a_i - b_i|$
Mahalanobis	$\sqrt{(a - b)^T S^{-1}(a - b)}$

3. Methodology

The analysis used the results of the survey conducted by IPSOS and Deloitte on the use of ICT in education in 2019. This survey aimed to identify information on use, access and attitudes towards ICT use of students, parents and teachers in 31 European countries. The survey was conducted in primary schools (ISCED 1), lower secondary schools (ISCED 2) and upper secondary schools (ISCED 3). The survey contained different themes: access and use of ICT, digital activities carried out by students and teachers, teaching skills and competences, use of ICT at home, digital policies and strategies. The survey was applied to a representative sample of 8,749 educational institutions. Table 2 presents a summary of the variables:

Table 2. Variables and Topics

Variable	Topic
Country_full	
Country_ID	
Language_country in 2 digits	
European Union	General characterization
School Identification	
Principal Identification	
Wave	
1 - Level	
2 - Level	
3A - Level	International educational level
3B - Level	

Participating principals	Number
Participating teachers	
Participating students	
Participating parents	
Boys at ISCED1	
Girls at ISCED1	
Boys at ISCED2	
Girls at ISCED2	
Boys at ISCED3 gen	
Girls at ISCED3 gen	
Boys at ISCED3 voc	
Girls at ISCED3 voc	
FTE teachers	
School community	
ADSL	Main means of internet access
Cable	
Fibre optic	
Wireless LAN	
Satellite	
None of these	
Don`t know/prefer not to say	
Broadband speed at school	
Desktop computer [ISCED 1 level]	Device used
Desktop computer [ISCED 2 level]	
Desktop computer [ISCED 3 academic level]	
Desktop computer [ISCED 3 vocational level]	
Desktop computer [All levels]	
Laptop or notebook [ISCED 1 level]	
Laptop or notebook [ISCED 2 level]	
Laptop or notebook [ISCED 3 academic level]	
Laptop or notebook [ISCED 3 vocational level]	
Laptop, or notebook [All levels]	
Tablet without Internet access [ISCED 1 level]	
Tablet without Internet access [ISCED 2 level]	
Tablet without Internet access [ISCED 3 academic level]	
Tablet without Internet access [ISCED 3 vocational level]	
Tablet without Internet access [All levels]	
Tablet with Wi-Fi Internet access [ISCED 1 level]	
Tablet with Wi-Fi Internet access [ISCED 2 level]	
Tablet with Wi-Fi Internet access [ISCED 3 academic level]	
Tablet with Wi-Fi Internet access [ISCED 3 vocational level]	
Tablet with Wi-Fi Internet access [All Levels]	
Tablet with mobile Internet access [ISCED 1 level]	
Tablet with mobile Internet access [ISCED 2 level]	
Tablet with mobile Internet access [ISCED 3 academic level]	
Tablet with mobile Internet access [ISCED 3 vocational level]	
Tablet with mobile Internet access [All levels]	
Digital reader [ISCED 1 level]	
Digital reader [ISCED 2 level]	
Digital reader [ISCED 3 academic level]	
Digital reader [ISCED 3 vocational level]	
Digital reader [All Levels]	
Interactive whiteboard [ISCED 1 level]	
Interactive whiteboard [ISCED 2 level]	
Interactive whiteboard [ISCED 3 academic level]	
Interactive whiteboard [ISCED 3 vocational level]	
Interactive whiteboard [All Levels]	
Digital camera [ISCED 1 level]	
Digital camera [ISCED 2 level]	
Digital camera [ISCED 3 academic level]	
Digital camera [ISCED 3 vocational level]	
Digital camera [All Levels]	
Equipment fully operational	
Computer laboratories	Number of desktop computers
Classrooms	
School library	
Other locations that are accessible to students in the school	
Digitally supportive school Cluster Analysis 2 (Support)	Digitally Supportive School
Digitally supportive school Groups	
Old var: without SC07	Digitally Equipped Schools
New var: with SC07	

4. Data and Discussion

For the purposes of this paper, only information related to the number of desktop computers in laboratories, classrooms, school libraries and other locations that were accessible in the schools was used:

4.1 Application of K Means Clustering

This technique subdivide the data sets into a set of *k* groups, where *k* is the number of groups pre-specified (Kassambara, 2017). The next two R Cran 3.6.3 libraries were used:

```
library(cluster)
library(factoextra)
```

For determinate the optimal number of clusters was used the function:

```
fviz_nbclust(europe, kmeans, method = "gap_stat")
```

Figure 1 presents the result of the optimal number of clusters:
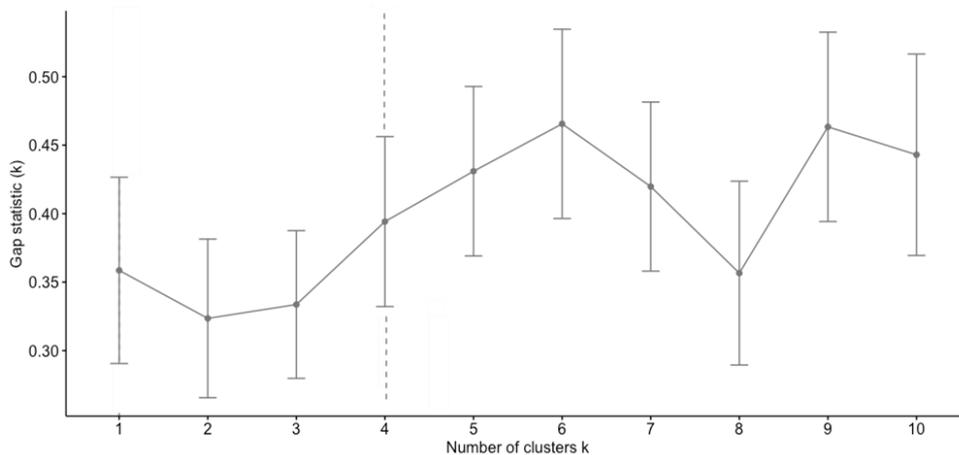
Figure 1. Clusters

Source: authors elaboration

The graph shows that the suggested number of clusters is 4. For determinate the countries in each cluster was used the function:

```
set.seed(123)  #  for  reproducibility  km.res  <-  kmeans(europe,  4,  nstart  =  25)  #  Visualize
fviz_cluster(km.res, data = mydata, palette = "jco", ggtheme = theme_minimal())
```

Figure 1 presents the cluster plot:
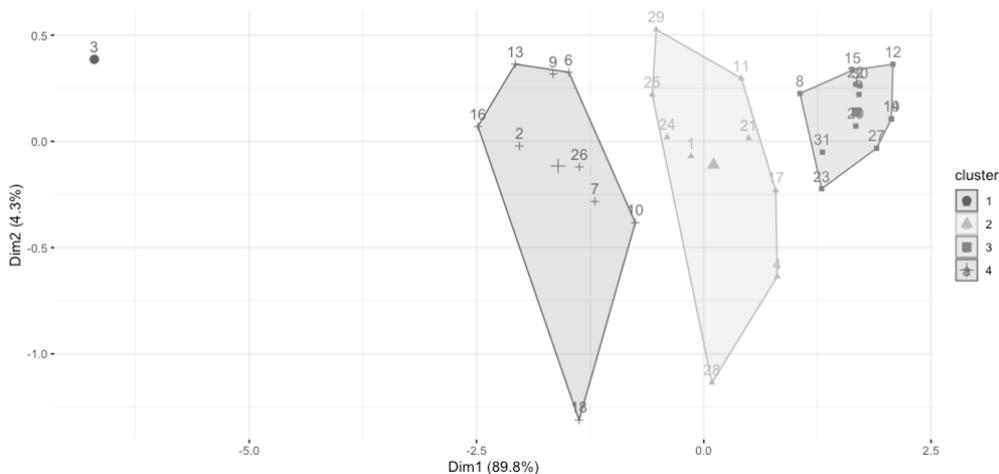
Figure 2. Clustering result

Source: authors elaboration

Cluster 1 (Bulgaria), *Cluster 2* (Croatia, Greece, Germany, Iceland, Belgium, Romania, Denmark and Spain), *Cluster 3*(Slovakia, Portugal, Poland, Austria, Finland, Malta, Italy, Cyprus and Slovenia) and *Cluster 4* (Estonia, UK, Norway, Ireland, France, Luxembourg, Sweden, Czech Republic, Hungary, Lithuania, Latvia, Netherlands and Turkey). Figure 3 presents the European political map:
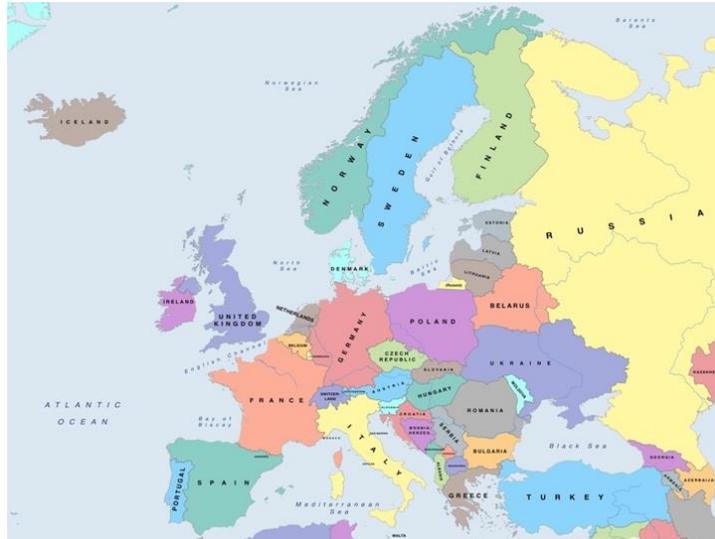
Figure 3. Political map of Europe

Source: (Agustin, 2018)

The *K-means* clustering results show that Bulgaria is the only country that was classified in an independent cluster. In effect, this country has the major number of desktop computers in laboratories, classrooms, school library and other locations that were accessible in the schools, which makes it possible to say that in this country investment in technologies for education is good. The second, third and fourth clusters grouped countries that are not together geographically but have a similar number of computers in the schools (laboratories, classrooms, school library and other locations). The four European countries with the major Gross Domestic Product (GDP) in 2019 were Germany, UK, France and Italy (Statistics Times, 2020), of these countries only UK and France were classified in the same cluster. All clusters have a different number of countries.

4.2 Application of Hierarchical Clustering

This technique doesn t require to pre-specify the number of clusters to be generated and the result of is a dendrogram (Kassambara, 2017). The dendrogram is the graphical representation of the clustering. Usually, it is drawn backward, starting from the final cluster with all the objects and from similarity 0 (Forina et al., 2002). The functions hclust(), dist() and fviz_dend() were used.

```
res.hc <- hclust(dist(europe), method = "ward.D2") fviz_dend(res.hc, cex = 0.5, k = 4, palette = "jco")
```

Figure 4 presents the cluster dendrogram:
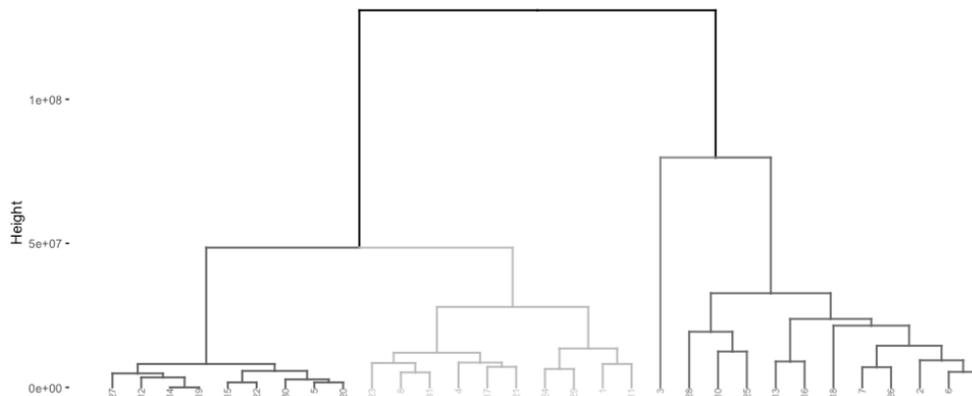
Figure 4. Dendrogram

Source: authors elaboration

Cluster 1 (Sweden, France, Hungary, Luxembourg, Ireland, Netherlands, Turkey, Czech Republic and Latvia), *Cluster 2* (Norway, Estonia, UK, Cyprus, Italy, Malta, Poland, Slovakia, Austria and Finland), *Cluster 3* (Bulgaria) and *Cluster 4* (Slovenia, Spain, Portugal, Croatia, Iceland, Lithuania, Denmark, Romania, Belgium, Germany and Greece). Similar to the *K-means* clustering, the hierarchical clustering results show that Bulgaria is the only country that was classified in an independent cluster and all clusters have a different number of countries. In this classification two (UK and Italy) of the four european countries with the major GDP were classified in the same cluster. Countries belonging to the same cluster have a similar number of desktop computers in their schools. Additionally, a heatmap was built. This is another way to visualize hierarchical clustering and consist in a graphical representation of data that uses a system of color-coding to represent different values (Lommatsch, Tucker, Moyer-Packenham, & Symanzik, 2018). The columns/rows of the data matrix are re-ordered according to the hierarchical clustering result, putting similar observations close to each other (DataNovia, 2018).

The R package *pheatmap* was used:

```
library(pheatmap) pheatmap(t(mydata), cutree_cols = 4)
```

Figure 5 presents the heatmap:
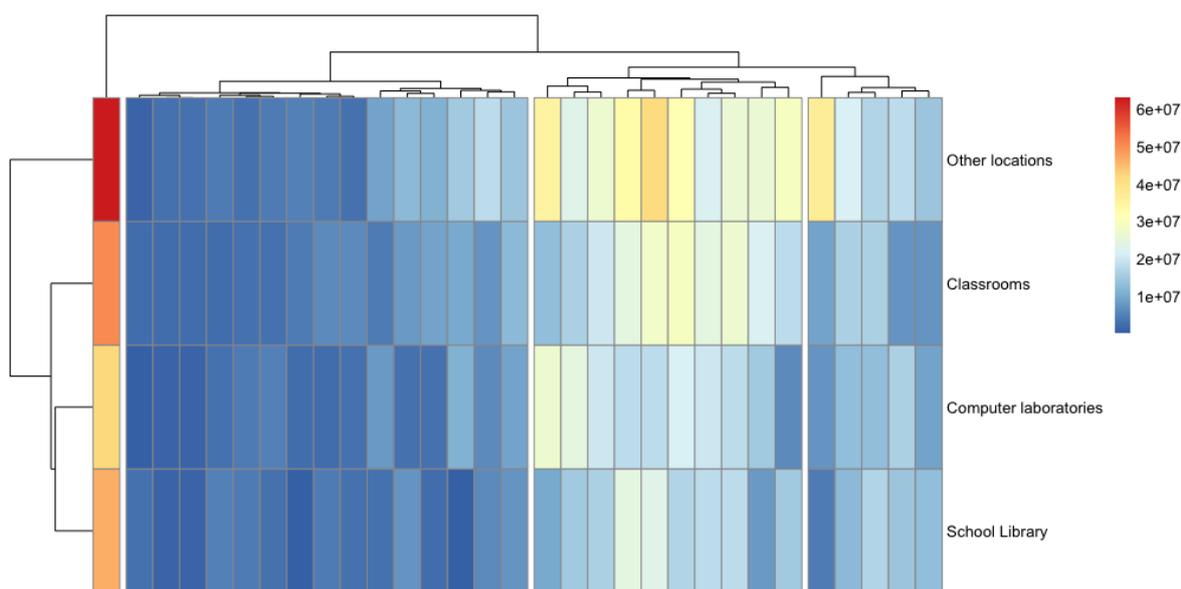
Figure 5. Heatmap

Source: authors elaboration

In each category (classrooms, laboratories, libraries and other locations), the blue, yellow, orange and red colors represent quantities. The dark blue color represents an amount close to 10.000.000 desktop computers and dark red color 60.000.000. This result shows that only one country has around 40.000.000 desktop computers in other locations (Bulgaria). In the majority of countries analyzed, the number of desktop computers in classrooms, laboratories and libraries was represented with blue color, i.e. the number in these places does not exceed 30.000.000 desktop computers.

5. Conclusions

The aim of this paper was to presents an application case of the clustering techniques, in particular the hierarchical and *k-means* clustering in the ICT use in European schools. Bulgaria is the country with the major number of desktop computers in their schools. The countries with the highest GDP are not the ones that have the most desktop computers in their schools, which may be due to the use of laptops. Overall, this paper intends to provide researchers with a brief guide to applied cluster analysis methods (*hierarchical* and *k-means*) to some dataset. The database contained sufficient information; the application of clustering techniques allowed obtaining valid information. Future research may use clustering techniques to analyze other aspects of ICT use in European schools.

In fact, Bulgarian schools provide conditions enabling the technology-enhanced teaching process, adequate to necessities of today's digital society. These include modern ICT infrastructure and resources, as well as opportunities for teachers to improve their ICT competence (Terzieva et al., 2019). The main reasons for the low number of desktop computers in schools in some European countries is that their use is concentrated mainly on the sporadic retrieval of information from the Internet. Only a small number of teachers use computer-assisted teaching and learning materials

regularly, as they report difficulties in integrating them into classroom practice, problems in allocating time for training and low ICT knowledge and skills. Furthermore, they lack the structural support as well as the infrastructure to find the most effective ways to apply ICT in the teaching process (Welzel et al., 2010).

References

Agustin. (2018). Europe Political Map Miller Large. Retrieved April 7, 2020, from Información Imágenes website: https://informacionimagenes.net/mapa-de-europa-con-nombres/europe-political-map-miller-large/

Arkorful, V., & Abaidoo, N. (2015). The role of e-learning, advantages and disadvantages of its adoption in higher education. *International Journal of Instructional Technology and Distance Learning*, *12*(1), 29–42.

Bhagat, A., Kshirsagar, N., Khodke, P., Dongre, K., & Ali, S. (2016). Penalty parameter selection for hierarchical data stream clustering. *Procedia Computer Science, 79*, 24-31. https://doi.org/10.1016/j.procs.2016.03.005

Bajunid, I. A. (2001). The Transformation of Malaysian Society through Technological Advantage: ICT and Education in Malaysia. *Journal of Southeast Asian Education*, *2*(1), 104–146.

Bouguettaya, A., Yu, Q., Liu, X., Zhou, X., & Song, A. (2015). Efficient agglomerative hierarchical clustering. *Expert Systems with Applications*, *42*(5), 2785–2797. https://doi.org/10.1016/j.eswa.2014.09.054

DataNovia. (2018). Heatmap in R: Static and Interactive Visualization. Retrieved April 7, 2020, from https://www.datanovia.com/en/lessons/heatmap-in-r-static-and-interactive-visualization/

Batanero, J. M. F., Rebollo, M. M. R., & Rueda, M. M. (2019). Impact of ICT on students with high abilities. Bibliographic review (2008–2018). *Computers & Education*, *137*, 48-58. https://doi.org/https://doi.org/10.1016/j.compedu.2019.04.007

Forina, M., Armanino, C., & Raggio, V. (2002). Clustering with dendrograms on interpretation variables. *Analytica Chimica Acta*, *454*(1), 13–19. https://doi.org/10.1016/S0003-2670(01)01517-3

Gong, X., & Richman, M. B. (1995). On the application of cluster analysis to growing season precipitation data in North America east of the Rockies. *Journal of Climate*, *8*(4), 897–931. https://doi.org/10.1175/1520-0442(1995)008<0897:OTAOCA>2.0.CO;2

Govender, P., & Sivakumar, V. (2020). Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019). *Atmospheric Pollution Research*, *11*(1), 40–56. https://doi.org/10.1016/j.apr.2019.09.009

Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems*, *17*(2–3), 107–145. https://doi.org/10.1023/A:1012801612483

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *28*(1), 100–108. https://doi.org/10.2307/2346830

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM Comp Surveys*, *31*. https://doi.org/10.1145/331499.331504

Jaiswal, S. (2018). K-Means Clustering in R Tutorial. Retrieved April 3, 2020, from https://www.datacamp.com/community/tutorials/k-means-clustering-r

Kassambara, A. (2017). *Practical guide to cluster analysis in R : unsupervised machine learning* (1st ed.; STHDA, Ed.).

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis* (Vol. 344). John Wiley & Sons.

Lin, J. M.-C. (2008). ICT education: To integrate or not to integrate? *British Journal of Educational Technology*, *39*(6), 1121–1123. https://doi.org/10.1111/j.1467-8535.2008.00825.x

Lommatsch, C. W., Tucker, S. I., Moyer-Packenham, P. S., & Symanzik, J. (2018). Heatmap and hierarchical clustering analysis to highlight changes in young children's developmental progressions using virtual manipulative mathematics apps. In *Using mobile technologies in the teaching and learning of mathematics* (pp. 167–187). Springer. https://doi.org/10.1007/978-3-319-90179-4_10

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, *1*(14), 281–297. Oakland, CA, USA.

Palomino, M. D. C. P. (2017). Teacher training in the use of ICT for inclusion: differences between early childhood and primary education. *Procedia-Social and Behavioral Sciences*, *237*, 144-149. https://doi.org/10.1016/j.sbspro.2017.02.055

Pathak, M. (2018). Hierarchical Clustering in R. Retrieved April 3, 2020, from Data Camp website: https://www.datacamp.com/community/tutorials/hierarchical-clustering-R

Sas. (2009). Proximity Measures. Retrieved April 7, 2020, from SAS/STAT(R) 9.2 User's Guide, Second Edition website: https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_distance_sect016.htm

Statistics Times. (2020). List of European countries by GDP. Retrieved April 7, 2020, from International Monetary Fund World Economic Outlook website: http://statisticstimes.com/economy/european-countries-by-gdp.php

Talebian, S., Mohammadi, H. M., & Rezvanfar, A. (2014). Information and communication technology (ICT) in higher education: advantages, disadvantages, conveniences and limitations of applying e-learning to agricultural students in Iran. *Procedia-Social and Behavioral Sciences*, *152*, 300–305. https://doi.org/10.1016/j.sbspro.2014.09.199

Terzieva, V., Paunova-Hubenova, E., Dimitrov, S., & Boneva, Y. (2019). ICT in STEM Education in Bulgaria. Advances in Intelligent Systems and Computing, 801–812. https://doi.org/10.1007/978-3-030-11932-4_74

Valentín, A., Mateos, P. M., González-Tablas, M. M., Pérez, L., López, E., & García, I. (2013). Motivation and learning strategies in the use of ICTs among university students. *Computers & Education*, *61*, 52–58. https://doi.org/https://doi.org/10.1016/j.compedu.2012.09.008

Van Ryzin, M. J., & Roseth, C. J. (2019). Cooperative learning effects on peer relations and alcohol use in middle school. *Journal of Applied Developmental Psychology*, *64*(August), 101059. https://doi.org/10.1016/j.appdev.2019.101059

Vega-Hernández, M.-C., Patino-Alonso, M.-C., & Galindo-Villardón, M.-P. (2018). Multivariate characterization of university students using the ICT for learning. *Computers & Education*, *121*, 124–130. https://doi.org/https://doi.org/10.1016/j.compedu.2018.03.004

Welzel, M., Graf, S., Sanchez, E., Fontanieu, V., Stadler, H., Raykova, Z., & Ioannidis, G. (2010). Application of computer aided learning environments in schools of six European countries. Contemporary science education research: scientific literacy and social aspects of science, 317-326.

Xu, Q., Zhang, Q., Liu, J., & Luo, B. (2020). Efficient synthetical clustering validity indexes for hierarchical clustering. *Expert Systems with Applications*, *151*, 113367. https://doi.org/https://doi.org/10.1016/j.eswa.2020.113367