Updating research on Chernoff faces for school cartography

José Jesús Reyes Nuñez[1], Anita Rohonczi[1], Cristina E. Juliarena de Moretti[2], Ana María Garra[2], Carmen Alicia Rey[2], María V. Alves de Castro[2], Anabella S. Dibiase[2], Teresa A. Saint Pierre[2], Mariana A. Campos[2]

[1]Department of Cartography and Geoinformatics, Eötvös Loránd University, Budapest, Hungary

[2]Centro Argentino de Cartografía, Buenos Aires, Argentina

Abstract Present work includes a short introduction about the Chernoff faces, emphasizing the importance of works related to its' use in cartography. The possibilities of this method of representation in thematic maps, more specifically in school cartography (adapting the principle followed by Chernoff on pictograms) were studied within an innovative theoretical research. Some of the practical experiences acquired during the theoretical and practical teaching of this method for MSc students on Cartography at Eotvos Lorand University (Budapest, Hungary) are also included. The theoretical research was also tested in an international project counting with the participation of Argentine and Hungarian specialists, presenting some of the more characteristic results briefly. Its conclusions are a starting point to follow the research within an international project with specialists of the Vienna University of Technology, trying to find answers to the questions that remained open.

1 Short introduction to the Chernoff faces

This method for data representation was created by Hermann Chernoff (at present Professor Emeritus of Applied Mathematics, Department of Statistics at Harvard University) in 1973 (Figure 1). The essence of his method is the use of the features of a human face to represent different variables, changing the parameters that determine a feature according to the values of these variables.

His original idea was to create a multivariate symbol easy to be recognized by readers mainly interested on statistical analysis. In the article written to introduce the method (Chernoff 1973) he affirmed that up to 18 themes or variables can be represented at same time.

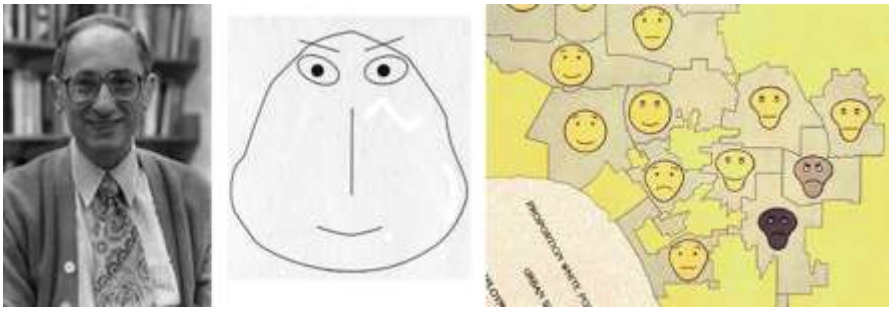

Fig. 1 Herman Chernoff's portrait, example of the first faces (1973) and a fragment of the first thematic map made by E. Turner (1977)

Only four years after the publication of this article, the method of representation began to be introduced also to cartography abroad, using the human faces to represent data on a map according to the traditional methods of thematic representation. The first and more famous (today considered a classic) example is the map entitled "Life in Los Angeles, 1970", designed by Eugene Turner and drafted by Richard Doss from the Geography Department at the California State University in 1977 (Figure 1). Turner wrote about this map: "*It is probably one of the most interesting maps I've created because the expressions evoke an emotional association with the data. Some people don't like that.*" (Turner 2004).

Other specialists from different scientific fields began to research the possibilities of the method beginning from the 90's. In this period the three more important names related to the correct use of Chernoff faces on maps are:

- Danny Dorling (University of Newcastle upon Tyne), who obtained his PhD degree on the theme of visualization of spatial structure, combining his cartograms with Chernoff faces to represent the results of data analysis about elections in Great Britain (Dorling 1991).
- Elizabeth S. Nelson, who beginning from the second half of the 90's was having detailed research on specific aspects as feature salience and natural correspondence on Chernoff faces, and the exam of search process using Chernoff faces (Nelson 1997–2007).
- Sarah I. Fabrikant (University of Zurich), who have developed numerous research on themes about data visualization, and in 2004 made a map entitled "Chernoff revisited: facing the presidential election" using morphed faces to represent the results of the elections in USA (Fabrikant 2004).

2 Theoretical analysis of the method and practical examples of proposed solutions for its use on maps

Originally, this method was not created for the data representation on maps. Herman Chernoff is not a cartographer or a graphic specialist. He proposed a

method for the graphic representation of data using a human face, but the characteristics of this method need to be adapted to the cartographic requirements before using it on a map. This is one of the reasons because the cartographic use of this method had not a real, more wide success during the past near 40 years (excepting some high quality maps created by cartographers and geographers, but they are a minority): some of the most important statistical software included the "map representation" of data visualized using Chernoff faces, but it was limited to draw very schematic, sometimes "caricaturistic" (nearly "antihuman") faces not on a map, but on a sketched representation of a territory (most times countries) delimited only by a very generalized (or a very roughed) borderline.

The schematic maps mentioned above reaffirm that the use of this method on maps requires a cartographer's experience to adapt it for the cartographic conditions in interest of making a readable map with good graphical quality. In 1998 I met Chernoff faces for first time, in an international workshop run by Prof. Henry Castner (Greensboro, USA) during an international ICA Symposium organized in Wroclaw (Poland). The exercises presented by Prof. Castner raised my interest in this topic, following the study of the method during the next years. The preliminary study of research and practical works related to the published Chernoff maps, and the detailed theoretical aspects of the present research were presented in the 24th International Cartographic Conference by the author (Reyes 2009). One main conclusion of this research was the limitation of the number of variables that can be represented by using an easy Chernoff face to a max. of six, applying some principles used successfully in cartography and described by Jacques Bertin in his Graphic Semiology (use of fill and change of size) in 1969 (Figure 2).

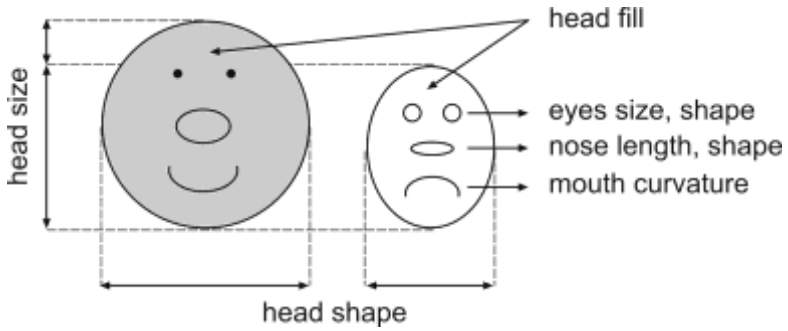

Fig. 2 The six graphic parameters proposed to change in a Chernoff face for thematic mapping (Reyes 2009)

The reason of this decision can be explained shortly: our main interest was to study the possibilities of the method in school cartography, so we had to make the face easy to be read by pupils from Elementary and Secondary Schools. In other hand, considering the human abilities to recognize graphic differences and to make graphic comparisons, the reading and analysis of numerous variables represented on a face (the maximal number was originally fixed on 18) is a very diffi-

cult task that demands too much time and attention. Studying some of the "maps" made with this method (and some variants of the method making it more detailed and growing the number of variables that can be represented), we can see that the time required to read and compare the information became as long as the read of the original statistical database. In those cases the graphic representation of the information does not fill its original objective of facilitating a faster and easier reading of the data.

In 2005 this topic was introduced as one of the themes to teach in the final semester of the subject entitled "Thematic Cartography" developed for MSc students of Cartography at Eötvös Loránd University in Budapest, Hungary. During the first years, students worked on creating their own thematic maps using Chernoff faces to represent four variables belonging to a main theme, trying to improve the faces with the use of graphic solutions developed by the traditional thematic cartography and combining the faces with other traditional methods of representation (e.g. choroplets) or using them as proportional symbols too.

During the practices numerous positive and negative experiences were acquired and discussed by the author with his students. In these discussions was proposed a possible new direction for a future research, which can be resumed with a short question: Why should only human faces to be used? Chernoff created more than a method of representation using a face: he determined a principle to divide a graphic symbol into its more relevant features or components, using each of these components to represent a different variable. Why not to apply this principle on map symbols, improving their traditional use in cartography?

For centuries symbols were mainly used to represent only one theme in the traditional cartography: more often the size (less frequently the shape, the fill or the outline) was changed to represent a theme. Using the principle followed by Chernoff to create the faces, new parameters can be changed within a symbol (Figure 3), and the number of themes to be represented would grow significantly.

This theory began to be experimented during the 2007/2008 school year, giving to the students a combined task: first of all to represent thematic data of selected Hungarian counties using Chernoff faces, and later to represent the same data on the same base map applying the Chernoff principle on cartographic symbols, more specifically on pictograms created by them.

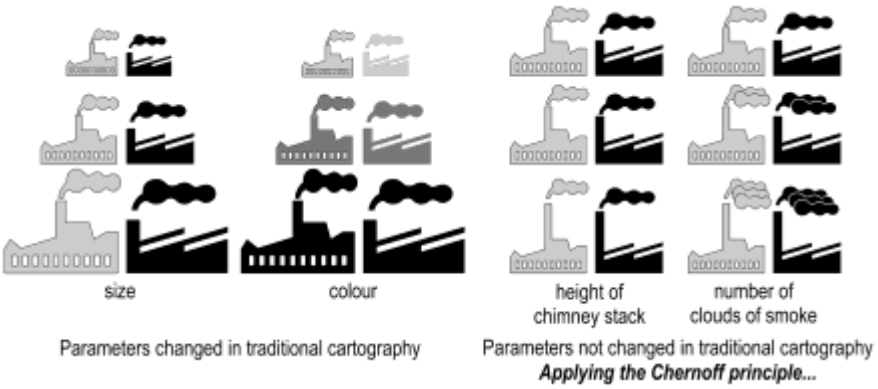

Fig. 3 Example for the use of the Chernoff principle on a cartographic symbol

Pictograms were selected because its components can be differentiated well, and depending on the theme and the design of the pictogram, it can be easier understandable for map readers than a geometric symbol. In other words:

- if a cartographer can choice a pictogram closely related to the main theme,
- and the features or components are clearly changed according the variables belonging to this theme.
- then the map reading should be easier and faster.

During the last two school years very interesting and original results were obtained with a high graphic level of realization. In the Figure 4 we can see fragments of different maps made by the students, working out themes as education, religion, ethnics, etc. In all the maps a general theme was represented using choroplets, constituting the background colour for the Chernoff faces and a compound pictogram was designed to represent other variables. Note that students used to "decorate" the faces with distinctive elements, in some cases using elements related to the topic, or to create a pictogram that "illustrate" the represented theme, trying to help and make easier the map reading.

Fig. 4 Fragments of maps made by Hungarian MSc students on Cartography

3 International project about the use of the method in schools

Our proposals related to the Chernoff method were tested during a representative survey in some selected Hungarian schools. Between 2004 and 2005 was organized a project entitled „Map reading by children in school age: Cartographic education and practice in Hungary and Argentina", developed under the scope of the bilateral agreement signed by both governments for the support of scientific research. The research on the possible uses of the Chernoff faces in thematic cartography with special attention to school cartography can be considered the continuation of the previously mentioned project, developed in two years (2008 and 2009).

3.1 Design of the questionnaire

Three main factors were considered during the previous organizative works of the survey:

- The specific characteristics of each educational system
- The real possibilities of the participant colleagues to organize the survey (Argentine specialists did not have any kind of financial support to execute the survey in their country, because the bilateral agreement finances only the exchange of specialists)
- The design of a questionnaire for pupils with some experience using maps and school atlases

Our final decision was to execute the survey for pupils of grades 7 and 8 in Hungarian Elementary Schools, and for pupils of 1st grade in Argentine Secondary Schools. Because of financial limitations the test was printed in a black and white A5 format. The detailed presentation of the survey and the questionnaire was made during the 3rd International Conference on Cartography and GIS (Nessebar, Bulgaria) in June of 2010 (Reyes et al. 2010).

After several consultations, the test was formed by four questions to examine four aspects of the use of Chernoff faces:

- Use of "traditional" Chernoff faces: the original method created by Chernoff has two main characteristics: only the shape of a face can be changed to represent data, and all the faces should be kept unfilled or filled with the same colour. The decision was taken to determine how difficult can be the reading of the data if only the shape was changed while the size of the faces remained the same.

- Use of Chernoff faces applying cartographic principles: the size of the faces was changed to represent a variable and in the Hungarian questionnaire was also changed the fill of the faces to represent a second one. These two parameters were not used by Chernoff in his original method, but they give us the opportunity to examine the grade of interrelation of the use of these parameters with the map reading (e.g. if the fill has or has no influence when the user reads other variables represented in the face). In the Hungarian questionnaire (Figure 5), the selected theme was the comparison of different agricultural products in the Hungary and other countries of the region (Austria, Czech Republic, Slovakia and Poland).

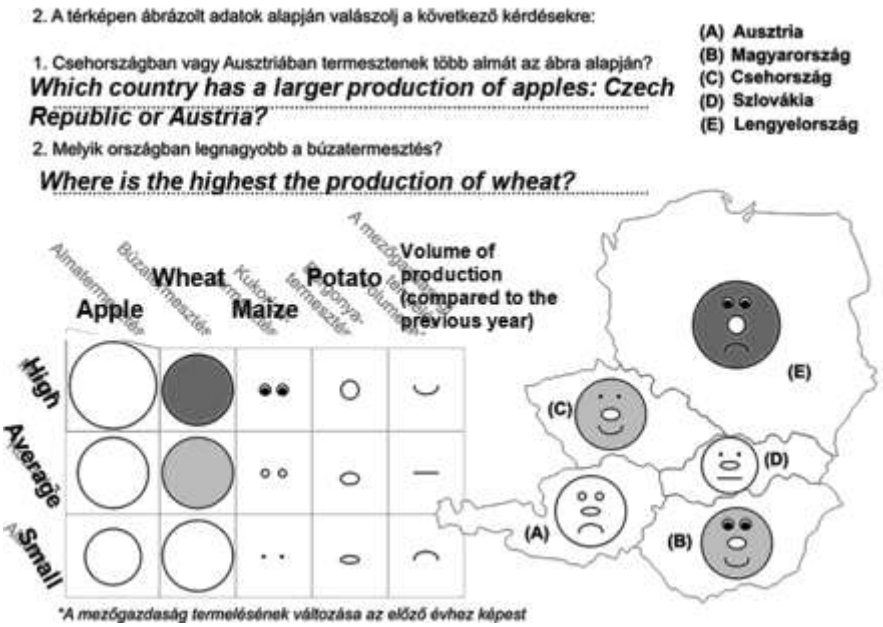

Fig. 5 Question presenting a more cartographic version of Chernoff faces in the Hungarian questionnaire (Reyes 2010)

- Applying the Chernoff principle on pictograms: This was one of the more interesting questions of the questionnaire, because the combination of the traditional cartographic pictograms and the principle used by Chernoff to create his faces was tested together (Figure 6). In Argentina and Hungary were selected two different themes (range of parks and squares in some districts of Buenos Aires, and the production of citrus in the southern provinces of Spain), but both themes were presented using the same pictogram (a tree) and using similar elements to represent the different themes. In the Argentine questionnaire (left side of Figure 6), the leafage was used to represent the total area of parks, the trunk represented the total area of larger squares, and the number of fruits represented the total area of smaller squares. Pupils were asked to identify districts with larger green areas, fewer parks and larger area for squares. In the Hungarian one (right side of Figure 6), the total production of citrus was represented with the leafage, and the production of oranges, mandarins and limes with the number of fruits, the number of boughs and the trunk respectively.

- Drawing of thematic data on an outline map using Chernoff faces: Pupils represented the data stored in a table using the preconceived legend.

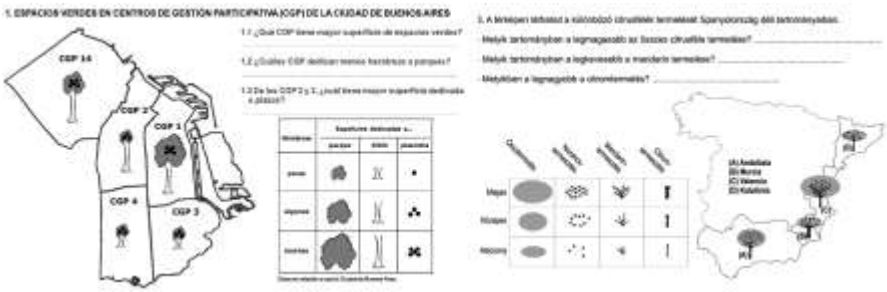

Fig. 6 Question applying the Chernoff principle on pictograms in the Argentine and Hungarian questionnaire

3.2 Survey in both countries

The questionnaire was applied between March and June of 2009. Argentine colleagues succeeded in collecting answers from 8 schools placed in the province of Buenos Aires. In Hungary a total of 12 schools participated from three provinces, arriving the majority of the answers from Budapest.

In Argentina a total of 818 pupils participated in the survey and the age group widely represented were the 13 years old pupils with 543 participants (Juliarena et al. 2009). A total of 1038 pupils answered the questions of the test in Hungary, and the majority of participants was constituted by 14 years old (437) and 13 years old (350) pupils (Figure 7).

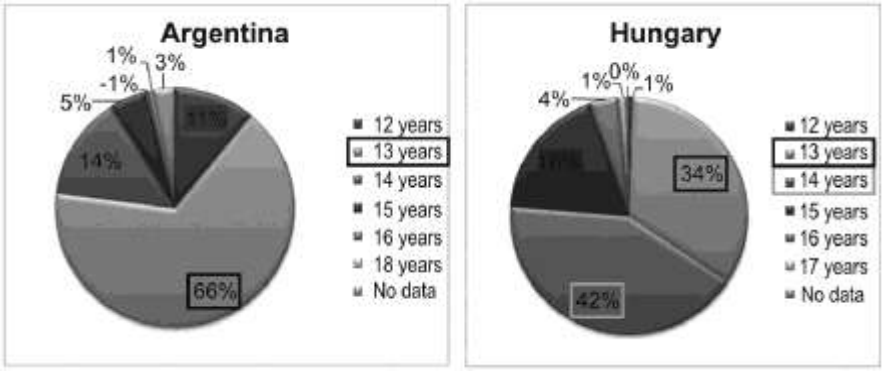

Fig. 7 Distribution of participants by age

3.3 General results and analysis of some answers

The general results of the survey are presented by questions and countries in table 1 and Figure 8.

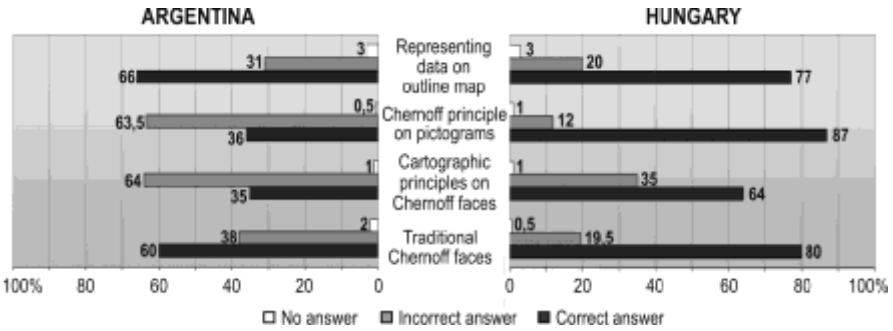

Fig. 8 Diagram comparing the percentage of correct and incorrect answers by questions in the survey

Together with the general results presented in the table 1, specialists also calculated the partial results by each question, which are presented by different diagrams on the website of the project (Reyes et al. 2009). Based on these diagrams we can have more genuine background information about the results of the survey.

MAIN RESULTS OF THE SURVEY						
QUESTIONNAIRE	**ARGENTINA**			**HUNGARY**		
Questions	**Right answers**	**Answers with one or more errors**	**No answer**	**Right answers**	**Answers with one or more errors**	**No answer**
"Traditional" Chernoff faces	493	313	12	828	207	3
Chernoff faces applying cartographic principles	285	527	6	665	367	6
Applying the Chernoff principle on pictograms	294	520	4	908	123	7
Drawing thematic data on an outline map with Chernoff faces	540	257	21	798	211	29

Table 1 General results of the survey

After consultations with the Argentine specialists, we should remark that the Argentine results reflect the need of the use of School Atlases during the teaching of Geography for pupils in Elementary and Secondary Schools, because this constitutes one of the reasons of the lower results of reading data represented in maps. At present, there is a lack of School Atlases in the Argentine Educational System,

and teachers and pupils use atlases not designed especially for the national curriculum on Geography. At the same time, the Geography textbooks used in Argentina contain more maps than the similar textbooks in other countries, trying to fill the absence of School Atlases. But as we could experience during our first research on reading thematic maps (Reyes et al. 2005), the textbooks cannot substitute the role to be played by School Atlases during teaching Geography.

The more contradictory result was obtained when the pupils gave answer to the question applying the Chernoff principle on a pictogram. As we can see in the table 1, the Argentine result was the second worse general result of the survey (correct answers were only a 36% of the total), while it was the best general result in the Hungarian one (87% of pupils answered correctly this question). In this case, the general result reflects a more negative situation than the analysis of partial results. After the analysis of the partial Argentine results that can be found on the Web (Reyes et al. 2009), it can be seen that the percentage of pupils with only one error in their answers (34%, 281 pupils) is very near to the 36% of pupils who answered correctly the whole question, and both categories together constitute nearly 70% of the participant pupils. Only a 7% of Argentine participants (52 pupils) did not give any correct answer to this question.

Other unforeseen result can be appreciated in the question of the Hungarian survey designing Chernoff faces applying also cartographic principles. In this case the size of the faces was changed to represent a data set (production of apples), and the fill was also changed depending on the production of wheat (Figure 5).

First we focused to compare the size between faces: *which country has a larger production of apples: Czech Republic or Austria?*, and the correct answer is *both countries*, because the size of the faces is the same. At same time the face representing Czech Republic was also filled with a darker grey tone (to represent a larger production of wheat) and it was also smiling (to symbolize a larger volume of agricultural production too). Our interest was to check if these attributes could have or not influence in the pupils' decision when they had to compare only the size of the faces. A total of 367 answers were wrong, and in 343 of them only one country was indicated (Czech Republic or Austria). In Figure 9 can be seen that only 112 pupils selected the Czech Republic over Austria, so we can affirm that neither the fill nor the smile represented an obstacle during their analysis to answer this question.

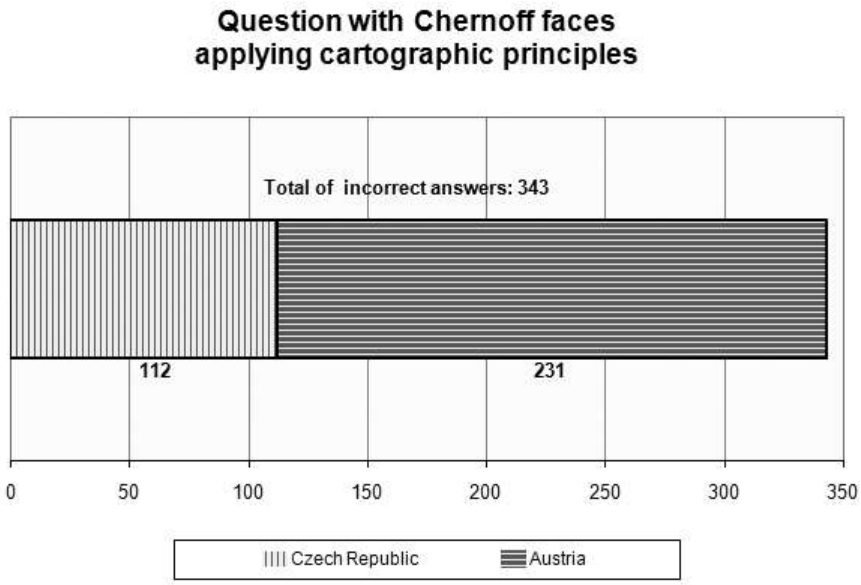

Fig. 9 Diagram presenting the percentage of answers to the question about the use Chernoff faces applying cartographic principles in Hungary

One of the more interesting results obtained in the Hungarian survey were the opinions given by 507 pupils (49% of the participants) about the method. Our first step for its analysis was to categorize the answers as positive or negative, and the result is presented in Figure 10.

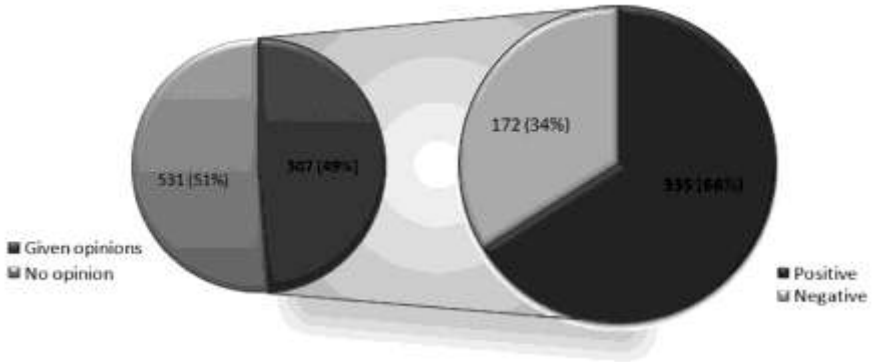

Fig. 10 Proportion of positive and negative opinions about the survey

Reading the answers, we can conclude that a considerable number of the negative opinions about the questionnaires came from pupils who resolved correctly all the questions. The relation of positive and negative answers by schools was also

examined and the results are represented in the diagram of the Figure 11 (using only a number for the identification of the schools to keep their anonymity).

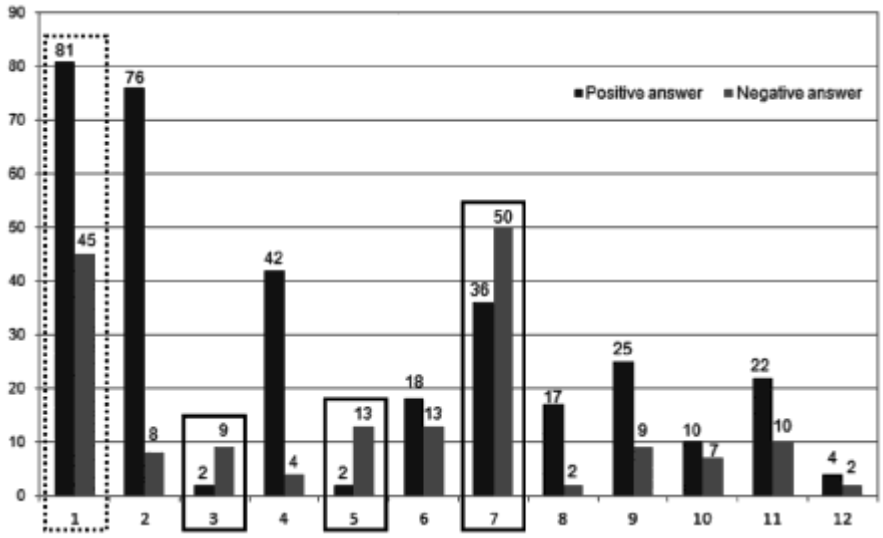

Fig. 11 Proportion of positive and negative opinions by schools

Four schools had results that emerged from the average: the larger number of negative opinions arrived from two schools (number 1 and 7) with a recognized high level of teaching in the country (but as we can see in the diagram, in the school 1 the number of positive opinions was higher than the negatives ones), while school number 5 can be considered of average level and school number 3 of below average. Mainly in the two first cases the common characteristics of their answers are the use of a polished vocabulary and the writing of the longer and more detailed arguments to express their opinions, sometimes combining negative and positive elements. Some examples:

- *Not a good idea: it is understandable and logical, but I do not see how help our thinking* (14 years old girl)
- *You need to pay higher attention, but more data can be plotted in less space. But I would not use it every day* (14 years old girl)
- *I like better the traditional symbols* (14 years old boy)
- *You need a longer time to read the legend, but it can help to endear the subject in earlier grades* (14 years old boy)

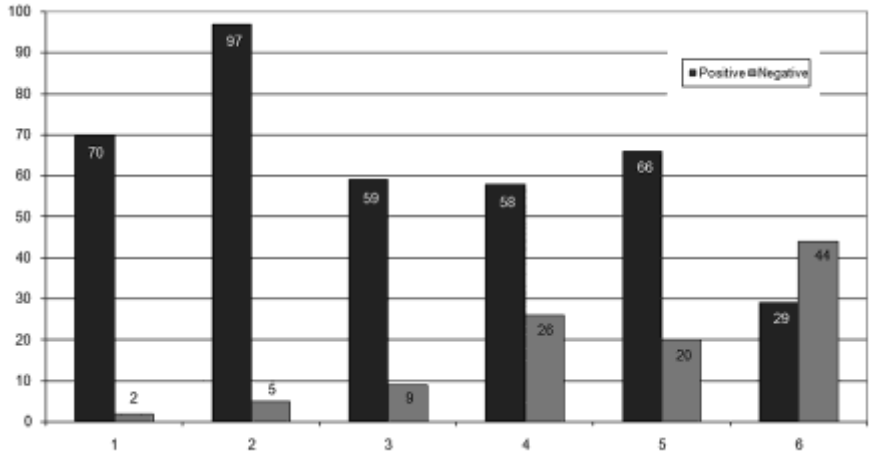

Fig. 12 Diagram showing the frequency of the numbered doublets in the pupils' opinions: (1) interesting – bored, (2) good – wrong, (3) like – do not like, (4) easy – hard, (5) funny – infantile (childlike), (6) understandable, suggestive, unequivocal – incomprehensible, inexplicable

We can affirm that a notable percentage of negative opinions are from pupils better prepared than the majority. This conclusion is also reaffirmed in the Figure 12, which represents some doublets with more frequent occurrence in the opinions. A total of six doublets were selected, expressing contradictory opinions like "interesting – bored", "easy – hard", etc. The summarized frequency of these words was compared as seen in the diagram. The doublet number 6 is the only one case, when the number of negative adjectives (incomprehensible, inexplicable) exceed the number of the positive ones (understandable, suggestive, unequivocal) and the selected vocabulary together with the expressed points of view let us infer that the 14 years pupils with a higher level of knowledge and using more often maps and school atlases in the classroom prefer to follow the use of the traditional methods of thematic representation (e.g. choroplet, diagram) for the visualization of data on maps.

At same time we cannot omit that they constitute a minority within the participant pupils: only a 34% of the total of opinions and barely a 17% of the total of Hungarian participants. Between the positive opinions (77% of the total) we can find some that directly or indirectly confirm the objectives set by us during the organization of the research and the survey:

- *I think our age group is accustomed to the traditional symbols, but this is a good idea for the smaller children* (13 years old girl)
- *I like this kind of symbolization, because it is more interesting than „coloring" and we can learn more of it* (13 years old boy)
- *Interesting, how many data can be represented with a face* (13 years old boy)

- *Interesting because many information is drawn in only one image* (15 years old girl)

-*I really liked the nature of these exercises, I would do it more times e.g. if more complex „shapes" are drawn to substitute the faces…* (14 years old boy)

- *Very good exercises, I like that the „image" is divided into several parts. Easy to understand, they could be used in more tests!* (15 years old boy)

Colleagues interested in this theme can find free access to all the databases, documents, etc related to this project visiting the following website: http://lazarus.elte.hu/hun/dolgozo/jesus/ma0809/proyect2.htm. All the documents are in two languages of the participant countries (Spanish and Hungarian),

4 Present and future plans

Some contradictory experiences can be noted when the Argentine and Hungarian results are compared (applying the Chernoff principle on pictograms, reading of data represented by changing only the shape of a face did not provoke more significant difficulties than if the representation is made changing the size and the fill). This survey was made for older (13-14 years old) pupils, but one of the conclusions of the previous theoretical research was that use of the faces can be more successfully for children in early grades of Elementary Schools. This idea was reaffirmed by some of the opinions given by the Hungarian pupils as can be read in the previous chapter.

Fig. 13 Website presenting the new project (fragment)

At same time that the periodical contacts with Argentine colleagues have been kept, expecting to begin new research in this and other themes, the Hungarian team also decided to follow the research in a new international project with the participation of the Institute of Cartography and Geoinformatics of the Vienna

University of Technology. According to our common decision, in 2010 both institutions began a project entitled “Further research and survey related to the theoretical and practical results of previous international projects about the possible cartographic uses of the Chernoff faces”, financed by the bilateral agreement for research between both countries (Figure 13). Our main aim is to find answers to the themes with contradictory results in the Argentine-Hungarian survey and to develop new research to complete these results.

Based on all the results, the specialists involved on this project plan to work out specific outlines about the possibilities of the use of the Chernoff method in school cartography, which can be useful for their future use in school atlases and other materials related to the geographical education.

References

Bertin J (1983) Semiology of graphics. University of Wisconsin Press

Chernoff H (1973) The use of faces to represent points in k-dimensional space graphically. Journal of the American Statistical Association, 68:361–367.

Dorling D (1991) The Visualization of Spatial Structure. PhD dissertation. Department of Geography, University of Newcastle upon Tyne.

http://www.sasi.group.shef.ac.uk/thesis/chapter8.html. Accessed at 9 January 2011

Fabrikant SI (2004) Blue and Red America http://www.geog.ucsb.edu/~sara/html/mapping/election/election04/election.html Accessed at 9 June 2010

Juliarena CE, Garra AM, Rey CA et al. (2009) Posible uso de las fases de Chernoff para la visualización de datos en la cartografía escolar. Boletin CAC 53(45) 2009/1:42–51

Nelson ES (2000) The Impact of Bivariate Symbol Design on Task Perfomance in a Map Setting. Cartographica, 37(4):61–78

Nelson ES (2007) The Face Symbol: Research Issues and Cartographic Potential. Cartographica 42(1):53–64

Nelson ES et al. (1997) Visual Search Processes and the Multivariate Point Symbol. Cartographica 34(4):19–33

Reyes JJ (2009) Ideas for the use of Chernoff faces in school cartography. The World's geo-spatial solutions. CD Proceedings of ICA 24th ICC. Santiago de Chile

Reyes JJ, Juliarena CE, Garra AM et al. (2005) Reading thematic maps in Argentine and Hungarian schools. Mapping Approaches into a Changing World. CD Proceedings of ICA 22nd ICC. A Coruna, Spain

Reyes JJ, Juliarena CE, Garra AM et al. (2009) Posibles usos de las fases de Chernoff para la visualización de datos en la cartografía escolar (2do año), http://lazarus.elte.hu/hun/dolgozo/jesus/ma0809/2/ekutatas.htm. Accessed at 30 December 2010

Reyes JJ, Juliarena CE, Garra AM et al. (2010) Chernoff survey in Argentine and Hungarian schools. CD Proceedings of the Third International Conference on Cartography and GIS. Nessebar, Bulgaria

Turner E (2004) Gene's Map Gallery http://www.csun.edu/~hfgeg005/eturner/gallery/gallery.htm Accessed at 9 January 2011

Acknowledgments The present research was developed in the framework of activities of the MTA-ELTE Research Group on Cartography and GIS. The theoretical research is supported by the European Union and co-financed by the European Social Fund (grant agreement no. TAMOP 4.2.1./B-09/1/KMR-2010-0003). The international surveys were supported by the project ARG-3/2007 of the National Office for Research and Technology of Hungary, and the project 68302 of the Hungarian Scientific Research Fund (OTKA).