Using Treemaps to represent medical data

Emmanuel CHAZARD[a,1], Philippe PUECH[a], Marc GREGOIRE[a], Régis BEUSCART[a]
[a] *Lille University Hospital, EA 2694, France*

Abstract. Confronted with the inadequacy of traditional charts, we tested the contribution of Treemaps to the representation of medical data. Treemap charts allow description of large hierarchical collections of quantitative data, on a synthetic way. Treemaps were implemented using PHP5, and were tested in the field of DRG-mining and other medical informations. From now on, this implementation is used in an interactive web-based request tool, and could be used to design interactive piloting tools.
Keywords : Medical Informatics, Computer Graphics, Treemaps, Charts

1. Introduction

Histograms, pie charts, bar diagrams and box plots are traditionally used to represent medical data. They are employed in a complex way, coupled with interminable tables, in order to render an account of the data's multidimensional structure. They show their limits with respect to the features being represented : several variables to be represented jointly, numerous classes among which the least populated cause information jamming, class overlap, very long wordings. Lastly, these frequently used tools do not exploit the interaction power that computers allow. Indeed, most end-users wish to see on the computer screen the same thing they used to read on paper sheets.

We would like to represent complex data by using areas, colors, and text labels.

2. About Treemaps

2.1. Introduction to Treemaps

Treemaps are a space-filling visualization method that allows to represent large hierarchical collections of quantitative data [1]. The principle consists in dividing the display area into a nested sequence of rectangles whose areas correspond to a quantitative variable. Treemaps have been applied to visualize files on a hard drive, as in Spacemonger [2], and to a wide variety of domains ranging from financial analysis [3], [4] to sports reporting [5]. Several distinct algorithms exist : the slice-and-dice [1], the cluster, and finally the squarrified algorithms. The squarrified algorithm [6] cuts out the available area so as to minimize the ratio of each rectangle. Thus, quantities are more easily comparable and the output looks more aesthetic. Treemaps seem to be one of the best ways to represent large hierarchical collections of quantitative data [7].

[1] Emmanuel Chazard - département d'information médicale, CHRU Lille, 2 avenue Oscar Lambret 59000 Lille France - emmanuelchazard@yahoo.fr

2.2. Problems related to scale

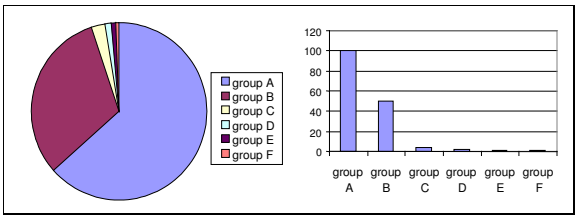

Figure 1 - ex. 1 - representing disparate quantities using usual charts

Although the usual graphs are drawn in two dimensions, they actually use only one dimension to represent quantities. As a matter of fact, in bar charts, the area is propotional to the sole height. Likewise, in pie charts, the area is proportional to the sole angle. Consequently, when a Cartesian scale is used, very disparate quantities induce reading confusion and small quantities become unreadable.

In order to illustrate this problem, we try to represent disparate quantities (going from 1 to 100), using traditional charts (Figure 1). The unidimensional charts induce difficulties of interpretation. Treemaps respond to this problem by drawing rectangles whose area is proportional to the quantity. On the Treemap chart (Figure 2), we can notice that group D is two times larger than both groups E and F.

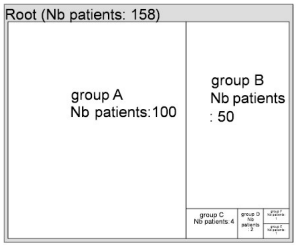

Figure 2 - ex. 1 - representing disparate quantities using Treemaps

However, the size of the rectangles is used to represent a volume (e.g. : "number of patients"), and not a density, (e.g. : "proportion of men").

2.3. Problems related to data overlap

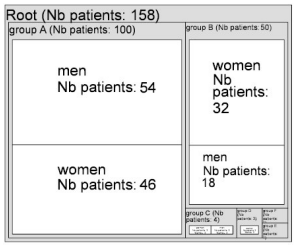

Figure 3 - ex. 2 - representing data overlap with Treemaps

The representation of data overlap is not very easy using conventional charts, and often results from a deliberate choice of the author (Figure 4). On the other hand, data overlap seems natural to Treemaps. As Treemaps consist in dividing a rectangle in several rectangles, those rectangles can be divided in their turn in a nested sequence (Figure 3).

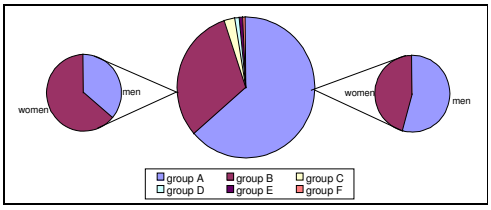

Figure 4 - ex. 2 - representing data overlap with pie charts

2.3.1. Problems related to color meaning

Contrary to traditional charts, Treemaps can represent an additional variable using a continuous color scale. Color scale should not be used to represent a variable sensitive to volume. As continuous color scales cannot be directly interpreted, the

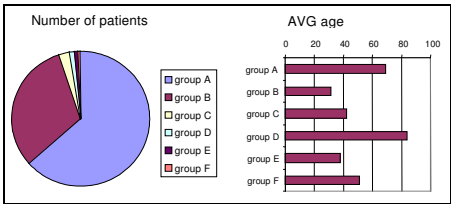

Figure 5 - ex. 3 - representing a second variable using traditional charts

values should appear in text labels.

In the next example, we try to represent the average age of the patient, in addition of the number of patients. Using conventional charts, we have to create a second chart (Figure 5). Using Treemaps, we can use colors (Figure 7). We can notice that the traditional charts would induce a misinterpretation of the global average age.

2.4. Problems related to data labelling

Traditional charts do not allow for the use of long text labels. Moreover, the size of text labels doesn't depend on the quantity represented, which induces diffculties in reading. In this example, we try to add long text labels to pie charts (Figure 6) and Treemaps (Figure 7). One can notice that values of the variables can be easily included in text labels.

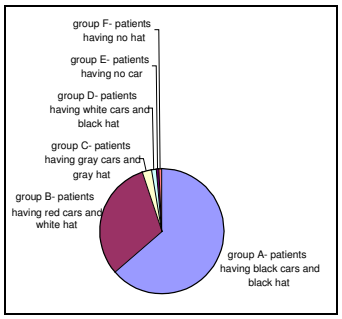

Figure 6 - ex. 4 - adding long text labels to pie charts

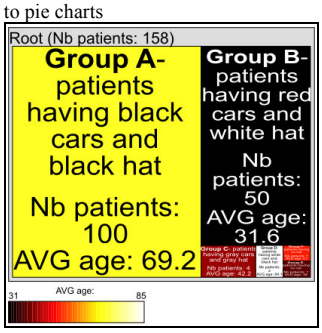

Figure 7 - ex. 3&4 - representing a variable with colors & adding long text labels to Treemaps

3. Material and methods

3.1. Program issues, output format

We wanted our program to be able to read hundred of files and to generate as many graphs on-the-fly, without any user interaction. Two output formats are available :

the JPEG format (Joint Photographic Experts Group) [8] :

- Advantages : VBA (Visual Basic for Applications) makes it possible to incorporate images of traditional formats within Microsoft Excel, such as BMP, GIF, JPEG.
- Disadvantages : no end-user interaction is possible ; the chart cannot be manually corrected ; poor quality printing

the SVG format (Scalable Vector Graphics) [9] : SVG is a XML-based format. The graph's description is written in formatted text according to XML standard. Then, a Web browser interpretes this description through a vectorial graph.

- Avantages : the end-user can zoom in and zoom out without any loss of quality ; the originator can allow end-user interaction, implementing onMouseOver effects, onClick effects, etc... [10] ; the imperfections of SVG graphs could be manually improved using some graphic softwares such as InkScape [11], or using a text editor such as Notepad, assuming SVG standards are known [9] ; high quality priting
- Disadvantages : the Web browser must be equipped with a plugin such as Adobe SVG Viewer [12], which was the case of all the hospital computers ; Unfortunately, VBA cannot load SVG graphics on the fly.

3.2. Color scale

We selected a "black-red-yellow-white" color scale (Figure 8) by respecting the following assertions :

- the scale is intuitive
- the colors can be ordered without any error
- the number of colours is rather weak, but the nuances have a strong discriminating capacity
- daltonians are not penalized
- printing in gray does not decrease legibility

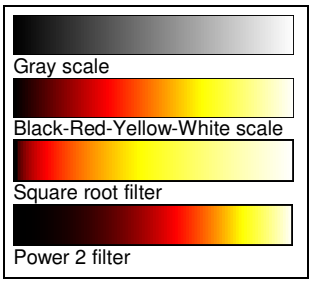

Figure 8 - color scales and filters

Canvas of increasing density could be used instead of colors, for better black & white document printing compliance. This solution cannot be adopted because of text labels. Our program also implements several nonCartesian filters.

3.3. Implementation

Several existing applications allow to generate Treemaps charts, like Microsoft Treemapper [13] (requires Microsoft Excel) or Treemap 4.1. [14] (requires Java Virtual Machine). Unfortunately, none of those applications could meet our needs.

We chose PHP5 (Php : Hypertext Preprocessor) [15] as a programming language. Its object-oriented structure [16] was particularly helpful to treat the rectangles overlap. PHP makes it possible to generate JPEG files by using the GD2 library [17], and to write text files, and thus to generate SVG, assuming SVG standards are known [9]. Our program implements the squarrified Treemap algorithm [6]. It consists in an object-oriented PHP class, which allows to reuse it in other coming programs.

4. Results

Table 1 - ex. 5 - extract from medical data : table

MDC	DRG	Nb Days	Daily Rate
MDC 40 Re-adjustment...	DRG 230 R.R. MDC - Age>=16 years – R.R. Care	201	238
MDC 40 Re-adjustment...	DRG 231 Age>=16 years – Handicap – phys dep<=12	752	176
MDC 40 Re-adjustment...	DRG 232 Age>=16 years – Handicap – phys dep>12	103	248

Our first exemple represents DRG related data. French rehabilitation care is financed by a prospective payment system, on the basis of daily rate [18]. Each day of the patient's stay is ranked into a Diagnosis-Related Group (DRG). DRGs are brought together into Major Diagnostic Categories (MDCs). A daily rate is linked to each DRG. We represent with both table (Table 1) and Treemap (Figure 10) the activity of a hole rehabilitation care hospital. Data are represented as following :

- each rectangle represents a DRG ; DRGs are brought together into MDC
- the area represents the number of invoiced days
- the color represents the daily rate of the DRG

Our second example shows geographical data. We want to represent the link between geographical origin and heavyness of the patients. Our territory is divided into 4 "life basin", themselves divided into 12 "health territory" (Figure 11). Data are represented as following (Figure 9) :

- each rectangle represents a health territory, brought together into life basins
- the area represents the number of patients comming from the health territory

- the color represents the proportion of heavy patients according to the Simplified Acute Physiologic Score [19]

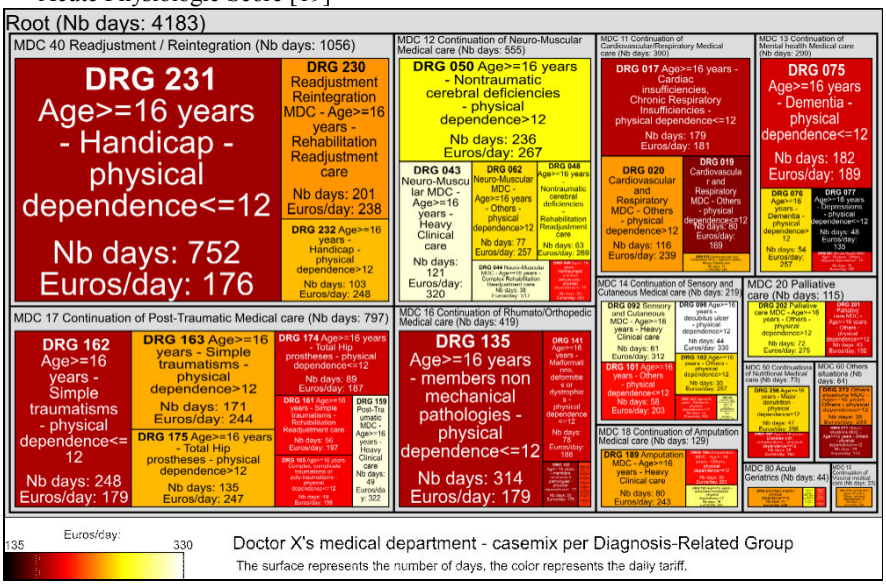

Figure 10 - ex. 5 - Complete output of the activity of a medical department, presenting simultaneously MDC, DRG, number of days, daily rate.

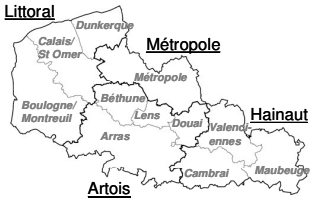

Figure 11 - ex. 6 - geographical splitting of the territory

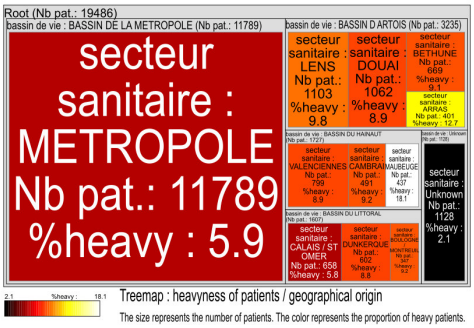

Figure 9 - ex. 6 - comparing number and heavyness of the patients with their geographical origin

5. Discussion

This implementation has several advantages. The use is simple and the result is immediate. Treemaps allow to represent a lot of information in the same chart ; the information is synthetic and treated on a hierarchical basis ; no additional table. They could be used to represent ratemaking data as well as geographical or epidemiological data. JPEG outputs allow to import charts into Excel generated dash-boards. SVG outputs allow interactive exploration and high quality printing.

However, certain weak points are to be taken into account. Treemaps are a “new” way to represent data and require a certain training. Outputs should be explained to the physicians. At the moment, treemap outputs cannot be obtained with traditional softwares. Treemaps are not usable for all kind of data. As written before, the size of the rectangles should not be used to represent density variables. They should not be

used to represent chronological data. Moreover, they can hardly be used to compare successive years, the rectangles being laid according to the current decreasing order.

Treemaps should be soon integrated into quarterly dash-boards. We also designed an interactive online request tool : from now on, end-users are able to choose interactively on a webform the variables to represent, from any network-linked computer. The output is generated on the fly, the end-user can explore interactively the output on the screen by simply using the mouse.

6. Conclusion

Although Treemaps are a synthetic and at the same time a very complete mode of representation, they are seldom used in industrial applications and never used in the health sector. Nowadays in France, new ratemaking rules induce a need for synthetic piloting tools, which could be fulfilled with Treemaps. Moreover, high-level interactive tool production should be an inventive way to motivate data providers.

Acknowlegements

Nicolas Leroy (EVALAB, University of Lille), Dr François Houyengah, Dr Omolade Alao, Vincent Chouraki, Dr Laurence Lefebvre, Thomas Ponchaut, Catherine Catteau (Department of Medical Information, Regional University Hospital of Lille).

References

[1] Shneiderman, B. Tree Visualization With Treemaps: A 2-D Space-Filling Approach. ACM Transactions on Graphics, 11(1), pp. 92-99.
[2] http://www.werkema.com/software/spacemonger.html (last accessed Jan. 12, 2006).
[3] Jungmeister WA, Turo D. Adapting treemaps to stock portfolio visualization. Tech Report CS-TR-2996, Computer Science Department, University of Maryland, College Park, MD.
[4] http://www.smartmoney.com/marketmap (last accessed Jan. 12, 2006).
[5] Jin L, Banks DC. TennisViewer: A Browser for Competition Trees. IEEE Computer Graphics and Applications, 17(4), pp. 63-65.
[6] Bruls M, Huizing K, van Wijk JJ. Squarified Treemaps. In Proceedings of Joint Eurographics and IEEE, TCVG Symposium on Visualization (TCVG 2000) IEEE Press, pp. 33-42.
[7] Bederson, B. B., Shneiderman, B., Wattenberg, M., Ordered and quantum treemaps: making effective use of 2D space to display hierarchies, ACM Transactions on Graphics, 21(4), Oct. 2002, 833-854.
[8] http://www.jpeg.org/ (last accessed Jan. 12, 2006).
[9] http://www.w3.org/Graphics/SVG/ (last accessed Jan. 12, 2006).
[10] Eisenberg DJ. SVG. O'Reilly, ISBN : 2-84177-228-4
[11] http://www.inkscape.org (last accessed Jan. 12, 2006).
[12] http://www.adobe.com/svg/viewer/install/main.html (last accessed Jan. 12, 2006).
[13] http://research.microsoft.com/community/treemapper/ (last accessed Jan. 12, 2006).
[14] http://www.cs.umd.edu/hcil/treemap (last accessed Jan. 12, 2006).
[15] http://www.php.net (last accessed Jan. 12, 2006).
[16] Gutmans A, Sæther Bakken S, Rethans D. PHP 5 Power Programming. Pearson Education, ISBN 0-131-47149-X
[17] Sklar D, Trachtenberg A. PHP en action. O'Reilly, ISBN: 2-84177-231-4
[18] http://www.atih.sante.fr (last accessed Jan. 12, 2006).
[19] Le Gall JR, Lemeshow S, Saulnier E. A New Simplified Acute Physiologic Score (SAPS 11) based on a European/North American multicenter study. JAMA, 1993; 270: 2957-63.