Visual Tools for the Exploration of Growth Data in a cohort of Kangaroo Infants during their first year of life

Deisy J. Díaz[1] ; Julieta Villegas[2]; John Alexis Guerra([1]; Nathalie Charpak[2]; José Tiberio Hernández[1]

[1]Grupo IMAGINE-Ingeniería de Sistemas y Computación, Universidad de los Andes;[2] Fundación Canguro

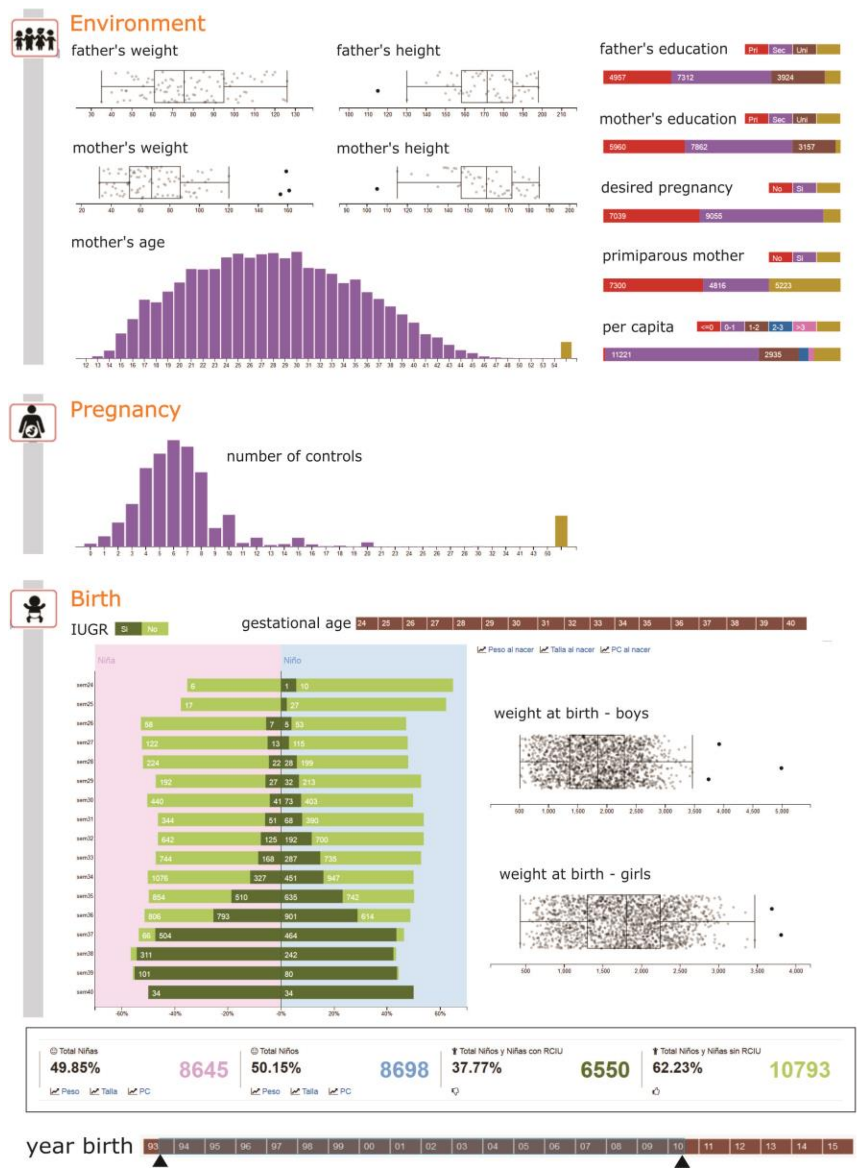

Figura 1: General view to explore 12-month follow-up data of Kangaroo Infants from 1993 to 2010

Abstract

The *Fundación Canguro* has been providing care for newborn premature infants in Colombia since 1994, helping them overcome the many issues related to their birth conditions. As part of their mission they have collected a longitudinal database of more than 33,000 children, recording their evolution during 23 years, with 47 measured variables. Having such rich data is both a blessing and a curse, as deriving insights from it is a daunting task. In this paper we present the results of a visual analytics

project that has been enabling the analysts of the *Fundación Canguro* to derive insights in ways they couldn't do before. To support the project, we present a process model that can be applied to any time based process where the results of an event are evaluated. For the problem at hand, the model considers the environment (before the event), pregnancy (development of the event), birth (the event), and growth (after the event). Using this model, a visual analytics tool was developed and implemented, and has been used by the analysts for the last months for deriving insights. The tool was developed applying Munzner's Visual Analytics framework [14] using advanced interactions and information visualizations, including a customized parallel coordinates view that also incorporate a time component. To demonstrate the value of the tool, we present two example case studies were analysts were able to conclude that educated women are less prone to have unwanted pregnancies, and that the implementation of a national law to increase maternity leave generated better feeding patterns on premature children in the Country.

INDEX TERMS: premature infant, visual analytics, parallel coordinates, Fundación Canguro, KMC method, Fenton curves, child growth

1 INTRODUCTION

Normal childbirth takes place between weeks 37 and 42 of gestation. When it occurs before week 37 it is considered a premature birth, which implies risks for the infants due to the fact that his/her organism is not completely developed and is not fully prepared for the new environment. The infant then has to face, possible infections as well as cardiac, respiratory, neurological, visual and digestive complications among others.

Globally, premature births are the first cause of infant mortality for children under 5 years. In Colombia, around 12% of child births are premature according to WHO. Prevention campaigns and information could reduce this rate which represents a heavy burden for both the health system and the families. [1]

The Fundación Canguro (Kangaroo Foundation) was created in Colombia in 1994 and has been helping worldwide premature and/or low birth weight (LBW) infants and their families through the Kangaroo Mother Care (KMC) Method [5]. As part of their mission the Foundation has been collecting a database of about 33.000 children, who have received care with the KMC. The database consists of 47 variables such as anthropometric measures (weight, height and head circumference) at birth and during the one year follow-up, data on the characteristics of the pregnancy and prenatal care, as well as the socioeconomic information of the parents. Extracting knowledge and deriving insights from such rich dataset that measures time-based effects of an event, especially one with these many variables, is a challenging task.

To address this task, we conducted an analysis tasks driven, user-centric project, which has enabled Fundación Canguro's analysts to derive insights from their data. It is of prime importance to highlight that this project was developed as a collaboration of technical as well as domain experts. As contributions of this project we present: 1) A model (which we believe can be generalized to other problem domains were the effects of an event across time), 2) A visual analytics tool that implements state of the art visual analytics techniques including a customized parallel coordinates view that adds a time component and 3) Two example case studies that illustrate the types of insights that the analysts were able to derive using our tool on their own.

The goal of our tool was to be able to evaluate the growth process during the first year of life of premature infants and to identify the influence of variables like the mother's age, education, weight, height or income in the process. To address this goal we propose a process model for problem domains when analysts want to evaluate the effects of a specific event, considering the before, during the development, the delivery and after the event. As illustrated on Figure 2, applying our contributed process model for the task at hand:

- The event or delivery is the birth of an LWB infant.
- The before consists of the socio-economic environment of the family at the moment of the pregnancy.
- The development encompasses the whole period of pregnancy.
- And, the after the growth variables measured on the infants.

Our process model can be easily applied to other problem domains where effects of an event are evaluated. As examples consider an electoral process with the candidates (before), the campaign (development), the election (delivery) and the performance of the winner (after), or a common factory production of goods with the materials (before), the actual fabrication (development), the product (delivery), and the ratings or acceptance of it (after). Given this generalizable model, we believe that our proposed visual analytics solution can be also applied to similar problem domains, although further research is required to validate this. More on the process model is illustrated on Section 2.

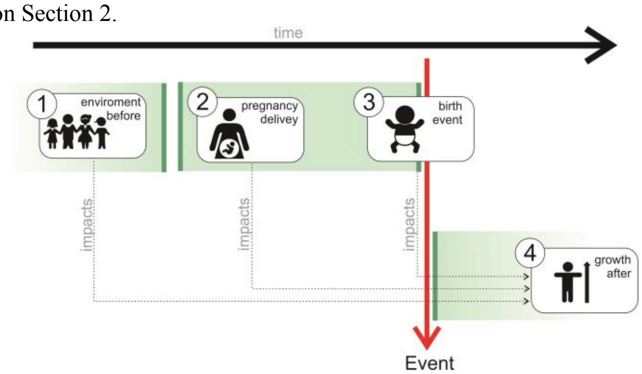

Figure 2: Variables were classified in these steps of the process

Once understood and modeled the problem. We developed a visual analytics tool applying Tamara Munzner's framework [14], which focuses on the abstraction of the What (the data), the Why (the tasks), and then the How (visualization idioms). As we describe the What consists of the 33,000 records categorized with our process model. The Why, focuses mainly on exploration of the variables, to support hypothesis testing of possible causes of the infant's growth. Finally for the how, we used coordinated views of four main components that reflect the four main components of our process model. Moreover we also contribute a customized parallel coordinates view that was enhanced to support the consideration of time with different levels of granularity. The proposed solution and this customized view are presented on Section 3.

The visual analytics tool has been used by analysts to derive insights that they couldn't do before, which demonstrate its success. To elucidate this we also provide two example case studies: one presented on Section 4, where analysts were able to check the effects of the mother's education on the levels of unwanted pregnancies, which uses the environment (*before*); and second described on Section 5, where they could identify the effects of a nationwide maternity leave law, on the feeding patterns evidenced on infants.

2 Proposed Model

2.1 Initial Information

The Fundación Canguro has gathered the information of around 33.000 infants that have been enrolled in the KMCP, keeping the register of their growth during the first year. The variables that have been defined are:

Socioeconomic variables: Per cápita income, education level of the parents and whether the pregnancy was desired

Prenatal variables: mother's age, height and weight of the parents.

Prenatal control variables: primiparous mother, number of pre-natal controls as a proxy of the quality of the prenatal care,.

Birth variables: gender, gestational age, intra-uterine growth restriction, prematurity.

Feeding patterns at 40 weeks of gestational age, 3, 6, 9, 12 months of corrected age (exclusive breastfeeding, formula or mixed feeding that includes breastfeeding and formula).

Anthropometric measurements (weight, height, head circumference) at birth, at entry in the KMCP, at 40 weeks of gestational age, at 3, 6, 9 and 12 months of corrected age.

In total there are 47 attributes with both categorical and quantitative information.

2.2 The process model

The *Fundación Canguro* not only wants to guarantee the wellbeing of premature infants through the implementation of the KMC method, but wishes to promote the quality of the KMC delivered and the development of preventive solutions such as empowering and educating mothers during the first year of follow up in the KMCP in the best nutrition strategy for their fragile infant.

In this context, a representation of the variables implied in each step of our process: environment, pregnancy, birth and follow up is offered showing existing relationship between them:

Environment: (Conditions *before*) the healthy growth of any child is influenced in great measure by the place and the family he/she is born into. The income and education level of the parents or even the presence of a brother and sister s can influence the later phases of the pregnancy, birth and growth of the baby. In this particular phase the variables included are those "relative to the environment": mother's age, mother's weight, mother's height, father's weight, father's height, mother's level of education, father's level of education, per capita income, whether the pregnancy was desired pregnancy, primiparous mother.

Pregnancy: (variables of *development*) corresponds to a transcendental step in the health of the newborn and will depend on the care the mother has during this phase. It implies that preventive measures must be taken and changes in lifestyles so as to include prenatal controls and medical exams to verify the growth and weight of the baby, as well as adopting healthy and balanced diets. For this step, the only variable considered is the number of controls as a proxy of the quality of prenatal care. The system, however, allows the entrance of any other variables when required or when new information is available, such as number of hospitalizations the mother required, diet, medicines provided during pregnancy, among others.

Birth: (Variables of *event*) birth is the core event in this process, and certain variables belonging to prior steps can affect the conditions in which the infant is born. The variables taken into account are: weight at birth, height at birth, head circumference (HC) at birth, intrauterine growth restriction (IUGR), gestational age in weeks.

Growth: (Variables of follow-up: *after*) for both infants born to term as well as those born prematurely, growth is an indicator of child health. There are several indicators that evaluate the nutritional state of a child. However the most frequent and simple to carry out in large groups are the anthropometric measurements, such as weight and height and head circumference that produce reliable information when adequately applied during the first year of life. The type of nutrition is fundamental to increase child survival rates and it has direct effect on growth: "if all children between 0 and 23 months were optimally breastfed, every year more than 800.000 children under 5 years, could be saved. Breastfeeding increases the intellectual quotient and school attendance, and it can be associated to better incomes in adult life." [11]

In the information available for this project, the anthropometric measurements (weight, height, head circumference) were taken during the first year of life as well as the type of nutrition that the infant received. Therefore the variables in this step are: weight, height HC, type of feeding at entry in the KMCP at 40 weeks (when reaching normal term), at month 3, 6, 9 and 12 of corrected age.

2.3 Method for Representing the Information

To represent each variable, different visual codes were taken into account (position, color, form, etc.) [13], or marks and channels as they are called by Tamara Munzner [14]. For example, the variable "mother's age" was represented by a bar graph (mark: line, channel: vertical position) which is a type of graph frequently employed by users as it is easy to understand.

Other more complex visualization were also created, like the graph used to represent the evolution of feeding types given to children at different points during their first year, which uses different marks: lines, circles and different channels like position, size and color in the same graph (Fig. 7)

3 Process Model Based Visual Analytics proposal Eneral Architecture

The software follows our process model presented before. We build a set of coordinated visualization for each step (Environment, Pregnancy, Birth, and Growth) in order to offer the free exploration of the variables following the analysis task defined by the user. We can observe them in Figure 1 for the first three steps.

The aim is to facilitate the flexible definition of sub-samples using the coordinated filtering features focused on user interest characteristics on *Environment, Pregnancy, and Birth*.

Based on each selected subsample, we offer our forth panel (*Growth*) to bring to the user an interactive analysis of growth follow up and feeding practices. (Fig. 3)

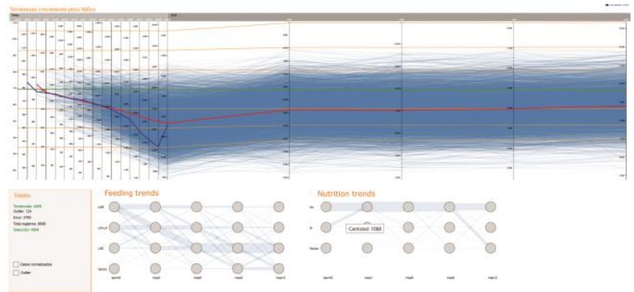

Figure 3: feeding trends for the selected sub sample

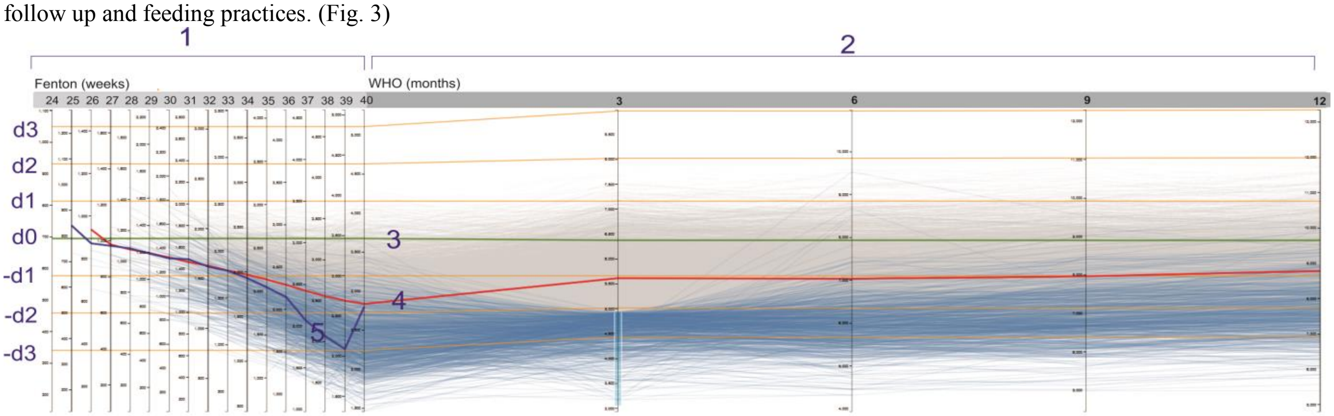

Figure 4: Growth trends representation (for a selected sample). We can see at the left, time in weeks (24th to 40th) and to the right, time in months (3,6,9,12). Red line is the mean of weight and bold Blue line is the mean of weight of babies born at this gestational age. Green line (d0) is the reference level (Fenton & WHO) and the levels of standard deviation upper and low are the di lines.

This panel allows selection, brushing, and filtering in addition the user can compare two or more subsamples.

3.1 A Contribution with parallel coordinates

Visualizing the evolution of a single child's growth over Fenton or WHO growth references curves [6] is useful and widely used in the pediatric area. The important aspect to consider is that growth of the infant should be consistent with the reference, without stall or jump in excess. When the growth is under -2SD (standard deviation), the diagnosis of malnutrition will be established.

However, to display the growth tendencies of a population sample simultaneously, graphs of parallel coordinates could be a good choice, given that they allow representations of different characteristics like: variability, tendencies, cluster, outliers. In this proposal, an adjustment is introduced to the classic technique o parallel coordinates to give the role to the position of the axes (dimension of baby's age) and thus visualize the evolution of an attribute of interest (weight, size or head circumference).

The objective is to take advantage of their potential for n-dimensional visualizations of the data. In this way the Cartesian coordinates of the Fenton or WHO curves are transformed in parallel coordinates allowing visualization relative to the expected tendency.

In contrast with the traditional diagram of parallel coordinates where each dimension corresponds to different variables that want to be compared or from which a correlation is looked for, in the proposed representation the dimensions correspond to a single variable: the age of the child, initiating in week 24 until week 40 and at 3, 6, 9 and 12 months of corrected age, that correspond to the precise points in time when the anthropometric measurements h are taken. Three graphs are created: one for weight, other for height and another for head circumference (HC).

The scales for each dimension correspond to the range of values that a child can have in weight, height or HC at the corresponding age.

With these types of visualizations, changes in growth during the first year of life of premature infants can be easily observed.

Nevertheless, in order to compare with the standard determined by Fenton (week 24 to week 40) or the WHO (month 3, 6, 9 and 12) the diagrams are complemented with the lines that represent the standard growth points according to Fenton and the WHO. The scale in each dimension is aligned so that the mean according to the standard in each week is set in the middle of the axis and the deviations -3, -2, -1 and 1, 2, 3 can be distributed upwards and

downwards respectively (Fig. 4).

For example in the case of the weight graph, on week 24 the scale goes from 500 to 1000 gr. and the mean value is 695.1 gr., while in month 12, the scale goes from 5700 to 12000 gr and the mean value is 9467.9 gr. If the growth of all the children were uniform from their birth to month 12, only horizontal lines should appear.

The user can observe easily the different trends of evolution of the weight of the infants in different periods of their first year. These changes can be due to factors such as feeding or illnesses they suffered, among other reasons.

3.2 Building the software: Analysis Task Driven-User centered process

Several models and frameworks to represent different types of data have been developed and proposed in VA projects that have solved particular situations where complex problems have to be solved. "The users, experts in their domains, are usually novices in VA subjects and could easily feel overwhelmed. What method can be used, how can I establish parameters or how can I move data from one place to another? Particularly when the visual analysis methods are not applied regularly, just occasionally, such questions are not easy to answer, a fact that makes VA difficult to use effectively in practice. Solutions that guide the user during the analysis and data exploration are what is needed."[16]

The process followed in the construction of this VA system is made up of the following phases.

3.2.1 Planning

In this phase, the general objectives of the project are analyzed with the user and a first set of analysis task is identified. An existing information analysis is performed. The sources of information are identified, data is gathered and a first understanding of the information is presented to the user. A first data quality interactive analysis is done with the user..

The results of this phase are automated processes to collect data, the definition of missing data treatment, standardization and storage processes for the information.

3.2.2 Exploration

In this phase the Dashboard and data visualizations with Visual Analytics tools are implemented aimed at satisfying the objectives of the project.

This set of tools becomes a first prototype which allows the user to familiarize with this type of software and the opportunity to solve questions formulated initially, to redefine questions or to generate new ones which were not taken into account, that are essential to the objectives of the project. It may not be the final tool, but is a first approach to what the user should expect of the system.

As a result of this phase, the following should be obtained: Initial exploration tools, answers to questions initially proposed in the objectives and identification of insights.

3.2.3 Guided exploration

This phase corresponds to the presentation to the user of the tools and the insights, and the steps followed to generate these insights. In this first exploration the user is guided while trying the tools. The outcome of this phase is: notes on the insights found together with the user, adjustments that have to be made to the exploration tool, statement of the first causality questions to determine specific analysis tasks to solve.

A new meeting is set to present a new exploration and to define the final analysis tasks.

3.2.4 Definition of specific analysis tasks

After defining and prioritizing the causality questions that have raised, and that are relevant to the user, the analysis tasks required to answer these questions have to be formally defined. This phase produces the development of Dashboard or *Gadgets* that help answer specific questions to enable the user to make decisions.

3.2.5 Recommendations

At this phase general recommendations with respect to the context of the problem are made together with the proposal of tools to continue with new visual analytics tasks.

4 USE CASE I: EDUCATION AND WANTED PREGNANCY

How does the variable "wanted pregnancy" behave when the mother's level of education is high (technical or university studies)?

The data exploration following this analysis task can be described as follow: By selecting mother's education level as *high*, in the graphic it can be seen that the variable *"wanted pregnancy*" the YES bar is bigger than the NO bar.

If we continue, maybe asking "how has this situation evolved through time?", then it can be observed that the behavior of the variable "*wanted pregnancy*" when the mother's level of education is *high*, is the same throughout observed time (comparing 1993-2009 with 2010-2015). However, if the selection of the variable mother's level of education is *low* (primary or secondary school), the YES bar and the NO bar of the "*wanted pregnancy*" variable are different form high education mothers (1993-2009), and different between the two observed time intervals!! (Fig. 5).

Before beginning to evaluate the child's growth, the expert has already discovered "insights" about the environment or the community to which the child belongs, and through these first explorations he/she can select a population sample with specific characteristics that he/she wants to analyze.

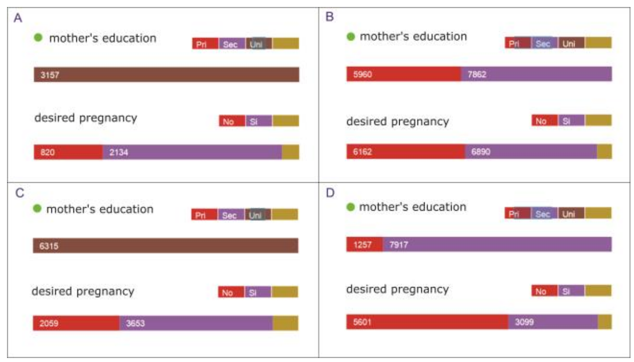

Figure 5: "Mother Education level" and "desired pregnancy" representation. We can observe a higher proportion of desired pregnancy in the high level education mothers.

5 USE CASE II: MATERNAL LEAVE AND FEEDING PATTERNS

Is there a change in the nutritional pattern of the infants after the nationwide law to increase the maternal leave introduced in 2011?

(i.e. increase in the rates of Exclusive Breastfeeding (EBF) at 3 months of corrected age, EBF at 40 weeks of gestational age)?

To answer this question, two types of visual representations were proposed: a graph of stepped bars for all the years, in each point (40 weeks – corrected age, 3, 6, 9, 12 months of age) with the different types of feeding alternatives: EBF, mixed(BF+F), exclusive artificial **f**ormulae milk (F). Changes of feeding alternatives that the infants have had during their first year, along 23 years of available information can be observed as an overview (Fig. 6).

The second visual representation is a graph that complements the first one and that shows the different tendencies and changes that occur in the infants's type of nutrition during their first year of life (Figs. 7).

The following conclusions could be made:

- From 1993 to 2010, the nutrition pattern that is highlighted is:

Week 40	EBF
Month 3	BF+F
Month 6	F
Month 9	F
Month 12	F

Although at week 40 the predominant type of nutrition is EBF in 1997, 1998 and 1999, the predominant type of nutrition is F at week 40.

- From 2011 to 2015 the nutrition pattern that is highlighted is:

Week 40	EBF
Month 3	BF+F
Month 6	BF+F
Month 9	BF+F
Month 12	BF+F

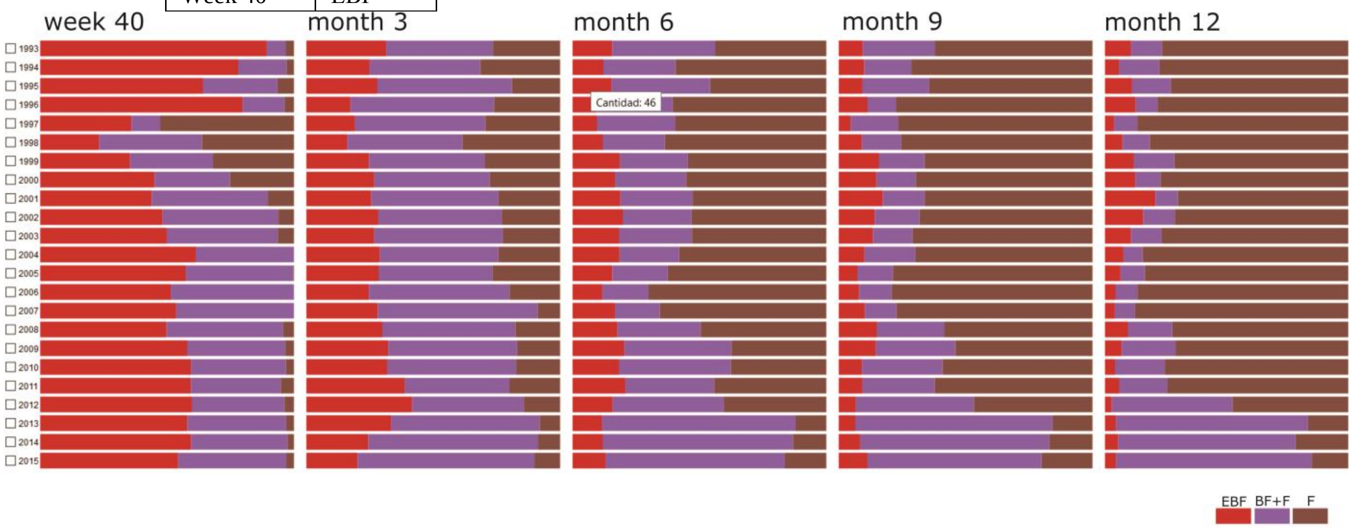

Figure 6: Representation of feeding type in the time, for the ages: week-40, month-3, month-6, and month 12.

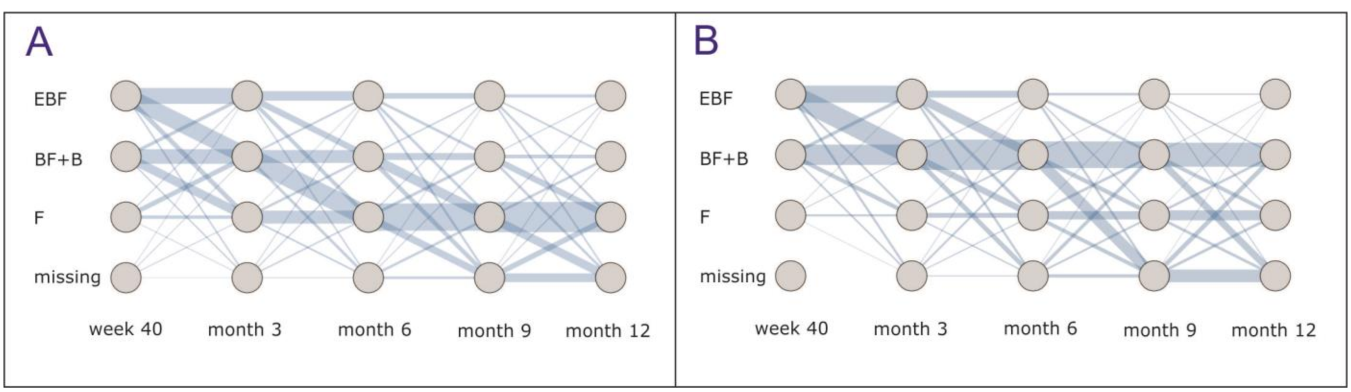

Figure 7: Use Case II. A: Feeding pattern from 1993 to 2010 B: Feeding pattern from 2011 to 2015

In month 3 the predominant food is the mixed BF+F, however an increase in EBF is remarkable in 2011 and 2012 compared to the rest of the years.

From 2011 onwards there is a decrease in F at each point.

In months 6,9,12 there was a considerable increase in the type of food BF+F with respect to the rest of the years, which means that the children´s food, F was replaced by the mixed feeding.

6 Conclusions

A set of interactive analysis tools focused on a 25 years long data on preterm infants was presented as well as the process to build it. A use case where experts use the tools and find insights on this interesting dataset was illustrated in order to show the main features of our proposal.

The use of different visualization techniques allows the observation of premature infants in many dimensions simultaneously, showing behaviors, patterns or trends, that can quickly to lead an hypothesis, or generate other new hypothesis, can also quickly check the status of the data.

The process model applied in our case (environment conditions, process-before, event, process-after) brings us a frame to guide the different analysis tasks with interactive visualizations. We hope, for future works, to apply it to other domains.

These types of interactive visualizations can be used to show if results of a policy achieve objectives or not. Additionally presenting the information in this way makes it easier for people to understand the situation of what is happening to be able to design policies agreed in this case for the benefit of premature infants

The proposed method (Analysis task driven, user centred) shows its benefits in terms of effectiveness in our support the analysis process. The collaborative process proposed and implemented is, from our point of view, a must in visual analytics tool development.

From technical point of view, the use of d3 as visualization library, j2ee and mongoDB on Heroku was a good combination with the proposed method. as the development gave the development an agile and robust environment without the need to add an additional cost to the Fundación Canguro.

7 Acknowledges

We would like to thank Fundacion Canguro for making accessible the dataset and the general support to this project.

References

[1] Howson, C. P., Kinney, M. V., McDougall, L., & Lawn, J. E. (2013). Born too soon: preterm birth matters. *Reproductive health*, *10*(1), S1.

[2] O. Bernal, C. Gutiérrez. (2012). Diagnostic and inequities in mother-child health in Colombia 1990-2010. (in Spanish) *Diagnóstico e inequidades de la salud materno-infantil en Colombia: 1990-2010*. Chap. 2 in: *La salud en Colombia Logros, retos y recomendaciones*. November 2012.

[3] PNUD. Millennium Development Goals Report Colombia 2015 (in Spanish) Available in: http://www.co.undp.org/content/dam/colombia/docs/ODM/undp-co-odsinformedoc-2015.pdf

[4] Oviedo, M. & García, M. C. (2011). El embarazo en situación de adolescencia: una impostura en la subjetividad femenina. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, 2 (9), pp. 929 - 943.

[5] Charpak, N., Gabriel Ruiz, J., Zupan, J., Cattaneo, A., Figueroa, Z., Tessier, R., ... & Mokhachane, M. (2005). Kangaroo mother care: 25 years after. *Acta Paediatrica*, *94*(5), 514-522.

[6] Fenton, T. R., & Kim, J. H. (2013). A systematic review and meta-analysis to revise the Fenton growth chart for preterm infants. *BMC pediatrics*, *13*(1), 59.

[7] Onis, M. (2006). WHO Child Growth Standards based on length/height, weight and age. *Acta paediatrica*, *95*(S450), 76-85.

[8] Funkquist, E.-L., Tuvemo, T., Jonsson, B., Serenius, F. and Nyqvist, K. (2010), Preterm appropriate for gestational age infants: size at birth explains subsequent growth. *Acta Pædiatrica*, 99: 1828–1833. doi:10.1111/j.1651-2227.2010.01966.x

[9] Fenton, Tanis R. A new growth chart for preterm babies: Babson and Benda's chart updated with recent data and a new format. *BMC Pediatrics*. Diciembre 2003. https://bmcpediatr.biomedcentral.com/articles/10.1186/1471-2431-3-13

[10] Fenton TR. Sauve R. Using the LMS method to calculate z-scores for the Fenton preterm infant growth chart. Eur J Clin Nutr. 2007;61:1380-5. http://www.nature.com/ejcn/journal/v61/n12/pdf/1602667a.pdf

[11] WHO (2016) Infant and young child feeding. Fact sheet. Available in: http://www.who.int/mediacentre/factsheets/fs342/en/

[12] UNICEF: (2012) Infants growth assessment (in Spanish. Evaluación del crecimiento de niños y niñas, julio de 2012. July 2012. ISBN: 978-92-806-4642-9

[13] W. S. Cleveland and R. McGill.(1984) Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association*, 79(387):531–554, 1984.

[14] Munzner, T. (2014). Visualization analysis and design. *CRC Press*. ISBN 9781466508910.

[15] B. Shneiderman. Dynamic queries for visual information seeking. *Software*, IEEE, 11(6):70–77, 1994.

[16] Ceneda, D., Miksch, S., Gschwandtner, T., Schulz, H. J., Streit, M., May, T., & Tominski, C. (2016) Characterizing Guidance in Visual Analytics, *IEEE Transactions on Visualization and Computer Graphics* doi: 10.1109/TVCG.2016.2598468