Case Studies for testing Insight Generation of BRAVIZ

Ana Mara Crdenas Gasca
Department of Systems Engineering and Computing
University of Los Andes
Bogota, Colombia
Email: am.cardenas926@uniandes.edu.co

Diego Angulo
Department of Systems Engineering and Computing
University of Los Andes
Bogota, Colombia
Email: da.angulo39@uniandes.edu.co

Jose Tiberio Hernandez
Department of Systems Engineering and Computing
University of Los Andes
Bogota, Colombia
Email: jhernand@uniandes.edu.co

***Abstract*—In the context of the development of a new tool for the exploration of neuroimage, clinical and socio-economic data (BRAVIZ), a method for evaluating its capabilities to stimulate insight was needed. Facing the difficulties of implementing traditional evaluation methods which required long periods of time, a large amount of expert users and the complications of defining quantitative metrics suited to insight generation; and based on a literature review of alternative evaluation methods suited to these restrictions; an appropriate and feasible methodology was designed. This methodology is based on an approach known as Case based studies that allows to review the experience of fewer users with no previous use of the tool with a qualitative approach. The evaluation methodology was applied in three sessions, documenting the experiences of 7 pairs of users working to solve their own questions around a provided data set. The conclusions obtained after the evaluation consisted in advantages and future improvements of the software as well as evidence about the insight generation and insight communication support provided by BRAVIZ.**

***Keywords*—Case Studies, Insight Generation Evaluation, Information Visualization , Exploratory Data Analysis, Neuroimaging**

I. Introduction

BRAVIZ was the tool being tested in this study. It is a collection of small applications for the exploration of neuroimage based data together with clinical and socio-economical data from each subject. It provides tools to visualize and extract scalar variables from structural, diffusion and functional images; and derived data as tractography, segmentation and surface reconstructions. Scalar variables can be visualized in parallel coordinates views, scatter plots, histograms and correlation matrices; and ANOVA and general linear models can be fit to them. All visualizations are linked, which permits connecting points in a statistical plot with the full spatial data associated with it.

The data used in this experiment was collected by the Kangaroo Foundation. The sample consisted of 450 young adults who were part of randomized control trial at birth. All of them were born prematurely or with very low weight, and were randomly assigned to the Kangaroo Mother Care treatment, or kept in the hospital for traditional care in incubator. These kids were followed during their eighteen months of life and all this data is available. Additionally 50 subjects born at term in the same hospital and from similar socioeconomic conditions were recruited. The subjects went through several neuropsychological tests (Intelligence, attention, verbal learning, motricity, mental health, etc); full pediatric examination, optometry and audiometry examination. Researchers also visited their house and place of study or work, and applied questionnaires to their parents and close friends. Records of their performance at school were recovered, and if they were working data about the type of job and the wage. Finally, 250 subjects born with less than 1800 grams and the reference sample from term subjects went through magnetic resonance imaging. The protocol included structural MPRAGE images, diffusion weighted imaging and 5 functional paradigms. Data was processed using Freesurfer, Camino and SPM to obtain automatic segmentations, surface reconstructions, tractography and functional statistical maps.

The main objective of the study was to determine whether or not the tool BRAVIZ could be used by its target audience to generate insight in the research of their respective fields. The study sought to determine how BRAVIZ was used as an advantage for resolving real research questions supported by neuroimage data taken from a real data set (KMC) by researchers from fields related to the data such as radiology, pediatry , neurology and psychology.

The main questions the study was looking to answer were:

- Is the tool successful in stimulating insights?
- How does the tool support the identification of data anomalies?
- Does the background of the users affect the performance in the use of the tool?
- How does the tool help the process of hypothesis generation?
- How does the tool help the communication of hypotheses and insights?

II. Related Work

Testing insight support capacities of a tool presents a challenge since no previous experience with evaluation methods regarding the objective of the study had been performed by the team members.Therefore, the proposed methodology is based on literature found on evaluation of insight support tools adjusted to the specific needs of this project.

Shneiderman and Plaisants paper regarding Multi-dimensional In-depth Long-term Case Studies[1] describe a methodology from which several points were taken into account : focusing on certain features, determining what would constitute professional success for the users, scheduling interviews and observation sessions, providing training and log tools, conducting interviews involving reflexions from the users and documenting achievements and needs of improvements.

Long-term case studies involves a study 1 to 3 years long but for the scope of this project this time was not feasible. The methodology also suggests modifying the tool under evaluation based on the results of each iteration before continuing the test.

The approach proposed by Perer and Shneidermans in the Case Studies of Gaining Clarity during Exploratory Data Analysis[2] is based on a methodology was executed with subjects who had already been working with their own data in their own projects. For the current study, since the tool was not yet being used by the public, the test methodology included both experienced and new users with the data set described in the introduction.

This methodology is divided into several stages considered useful for the evaluation of BRAVIZ: first, interviewing the subjects before the test to know their study interests and their research fields. Second, training the users on the tool and performing a final interview to understand the outcomes of the work. Because of time restrictions, unlike the second methodology, the data was the same for all subjects and the working sessions were not carried out individually in their own workplaces. However, there were no restrictions on the research questions.

A third methodology was reviewed from Duca, Saraiya and Norths presentation on an Evaluation of Microarray Visualization Tools for Biological Insight[3]. Although this work shared the objective of answering how successful was the tool in stimulating insight, the approach followed in this methodology is based on quantitative methods while a qualitative evaluation was more appropriate for BRAVIZ . Although the resulting method is distinct, it relays on the definition reached by the authors of insight as an *individual observation about the data by the participant* and a *a unit of discovery*. This definition of insight was characterized by a timing, an observation, significance, hypothesis and correctness. However, for the scope of BRAVIZ testing, correctness and timing were not as important as the finding itself.

Finally, the paper Seven Guiding Scenarios for Information Visualization Evaluation [4] provides a compendium of several methodologies and categories. The scenarios for the current study were identified as an evaluation of visual data analysis and reasoning and an evaluation of collaborative data analysis. The first one is defined as *Evaluations in the studying if and how a visualization tool supports users in generating actionable and relevant knowledge in their domain*. For this scenario the paper suggests 2 types of methodology: Case Studies like MILC [1] and Controlled Experiments [3].

As for the Evaluation of Communication through Visualization the paper states that *Communication can pertain to aspects such as learning, teaching, and idea presentation as well as casual consumption of visual information as in ambient displays* [4]. The methods suggested for this kind of evaluation were Controlled Experiments as a quantitative approximation and Field Observation and Interviews as qualitative methods.

III. Methodology Description

Based on the reviewed literature and considering time restrictions the chosen methodology was based on individual case studies of subjects, some of whom had work with the tool before. The users belong to several research fields supported by neuroimage based , clinical and socio-economical data. Because some of the researchers had never used the tool, there were training sessions before the working sessions as suggested by [1].

The methodology followed was divided in the following stages:

A. First Interview

This first interview was a mean to evaluate previous knowledge of the tool as well as understanding the actual research interests of each participant and their experience with other visualization tools.

The interview was guided. Questions would change depending on the subject but the main points to be clarified were:

- Field of study and research interests.
- Research projects involving analysis of data with support of visualization or data analysis tools.
- Experience with other visualization or data analysis tools.
- Difficulties faced when using visualization or data analysis tools.
- Experience with BRAVIZ.
- Knowledge of KMC data.

B. Session 1

The first session was used for training and introducing the KMC data set. It was programmed to last 1 hour. The data presentation included an introduction to the KMC project and the introduction of the neuroimage based data, clinical and socio-economical data from the project. The name of the variables and the explanation of some of the tests used for collecting the neuroimage based data were also presented.

After the data presentation a questionnaire was applied. In the questionnaire users were asked to formulate research questions and answer how would they solve these questions with their traditional workflow.

For this session, the training included an introduction to the features of the software and a hands-on session where the users could interact with the features guided by the instructor. The hands on session focused on certain features of the software such as: Subject Overview and the different brain visualizations of this feature; fMRI Explorer and data analysis tools such as Anova, Linear Model, Correlations and Parallel Coordinates. Also, questions about the software were answered.

Finally, after the hands-on, users were asked to outline their possible workflow to solve the questions they had formulated at the start of the session using BRAVIZ.

C. Session 2

The second session started with a short review of the data. Afterwards users were asked to work in pairs with BRAVIZ in one of the questions formulated at the previous session. It was programmed to last about 1 hour. Along the working session, guided interviews were conducted reviewing the current task performed and the difficulties faced. The process was documented recording audio and video of each participant with the consent of the users.

At the end of the session a general review was performed. Each group shared their discovered insights, their difficulties while using the tool as well as the advantages they found. A final general feedback of the tool was shared among all participants. This process was also documented with video.

IV. Experiment

A. Group 1:

Group Members:

- Dr. Jorge Humberto Marn: Dr. Marn is a specialist in Radiology and Diagnostic Imaging. His research includes toxicological, neuropsychological and neuroimaging correlation in psychoactive substances users. His research is conducted in the San Jose university hospital. He has been working with the Kangaroo project and had previous experience using the BRAVIZ software.
- Dr. Nathalie Charpak: Dr. Charpak is a Pediatrician and director of the Kangaroo Foundation. She is the user with more experience with BRAVIZ as well as with the KMC data set.

Questions Reviewed:

The question reviewed was *Is the volume of the vermis of the cerebellum different between Kangaroo and Incubator treated children, when controlling for gestational age and excluding babies born after 37 weeks of gestational age?*. Since both members of the group had previous experience with BRAVIZ and Dr. Charpak used the software as part of her usual workflow they started with the resolution of the question in BRAVIZ without any planning.

Working Session:

In the practical working session the group used a filter for the selection of preterm infants with good quality MRI images. To answer the question they asked, the group found itself in the need of segmenting the vermis volume and tried to use the ROI builder without success . After that they decided to use the whole volume of the cerebellum and perform ANOVA test with left and right cerebellum for white and grey matter.

Difficulties and results:

This group had difficulties since they could not find the variables for the specific volume needed. The ROI feature was not discussed in the hands on since it implied a more advanced knowledge of the application. They suggested that a ROI feature that could allow not only spheres but free-hand drawing would be more practical. Difficulties to create the volume were not only due to the feature but because building volumes requires a detailed knowledge of the anatomical factor and biomarkers.

B. Group 2:

Group Members:

- Dr. Charles Guttmann : Dr. Gruttmann is the Director of the Center for Neurological Imaging at Brigham and Women's Hospital and an Assistant Professor in radiology at Harvard Medical School. His main interest is the quantitative evaluation of normal and pathological states of the brain using MRI. Dr. Guttmann works on SPINE (Structured Planning and Implementation of Neurological Explorations) [4], which is an informatics infrastructure designed to enable topical consortia and collaborative research studies. He has worked with several data analysis and visualization applications before, such as 3D SLICER, R and Matlab-based image analysis tools.
- Pablo Reyes: Pablo Reyes is a Clinical neuropsychologist and researcher, with deep experience in the diagnosis and research of degenerative diseases like Alzheimer's and frontotemporal dementia. As for his experience with visualization, he has research projects using neuroimaging tools such as functional MRI, DTI, ASL (Arterial Spin Labeling) to explain cognitive models.

Questions Reviewed:

This group reviewed 3 questions *Is the volume of the vermis of the cerebellum different between Kangaroo and Incubator treated children, when controlling for gestational age and excluding babies born after 37 weeks of gestational age?*, *Is thalamic volume different between Kangaroo and Incubator babies at 20 years old?* and *Is MD in the corpus callosum different between Kangaroo and Incubator babies?*. When asked how they would approach this question with their traditional workflow the group answered that they would extract volume with Freesurfer, and use SPSS, EXCEL or QDEC (from Freesurfer).

Working Session:

In the working session this group created a subsample of subjects that were born before 37 weeks of gestational age. This subsample was used in an ANOVA test with the white matter volumes of the cerebellum (since the volume of the vermis was not available) as the outcome and whether they were kangaroo or not and ballard test result as regressors to answer the first question. After the ANOVA test the group created a new variable with the volumes of the left and right thalamus and performed a linear model in order to answer the second question. They did not have enough time to complete the third question.

Difficulties and results:

This group faced difficulties finding the name of the variables they needed in the dataset because some names were

in spanish and initially they were planning to work with the volume of the Vermis but this was not available. They found that some values had an inconvenient scale (very small values for mean diffusivity) and had problems determining whether or not the sample was already loaded in the application.

The insights obtained by this group were that belonging to the kangaroo group did not have a significant effect on the volume of the cerebellum; but a significant effect was found in the volume of the left thalamus with respect to being term or preterm.

C. Iteration 1 (Groups 1, 2)

In the final feedback doctor Guttman remarked on how the tool is *easy to play around with*, although he found missing some interactions with the visualization like how to tell if he can trust or not a given structure, to what degree a Freesurfer volume is accurate and the *reputation* for a given measure. Dr. Charpak noted that the visualization itself helps sometimes to know when the data is wrong.

Pablo Reyes noted that the interaction with the variables could be improved with a hierarchical structure or by using keywords instead of the full list of variables.

Finally, Dr. Marn remarked that sub segmentation presented difficulties since it required a long process to get the new measure validated, calling it a *matter of the biomarkers and anatomical knowledge and having the right knowledge of the region that is going to be sub segmented*.

D. Group 3

Group Members:

- Dr. Marn helped us again in the second session participating with some of his students.
- Dr. Juan Monsalve: Dr. Monsalves speciality is neurology and epidemiology. When interviewed about his experience with visualizations and data analysis software he declared that as an epidemiologist he used SPSS, Stata and Epi Info. He has also worked with an in house software from the hospital for Magnetic Resonance Imaging, Tractography and Diffusion.

Questions Reviewed:

This group wanted to review the relation between having or not to the kangaroo treatment, the intellectual coefficient and volumes of the brain cortex. When asked for an approach with their regular workflow the group mentioned the use of FreeSurfer and BRAVIZ since Dr. Marn had previous experience with this tool.

Working Session:

In the working session this group created samples for kangaroo and not kangaroo and performed an ANOVA test with the intellectual coefficient as outcome. This group did not had enough time in the session to relate the results obtained in the ANOVA test with visualizations.

Difficulties and results:

One of the objectives of the BRAVIZ evaluation was to determine if the tool was useful for detecting data anomalies. During the working session the group included in the ANOVA test an empty variable and therefore the test did not work. This was an indicator that although users can see the distribution of data while they select variables, it was not obvious for them that the variable contained no data.

E. Group 4

Group Members:

- Dr. July Vanessa Martnez : Dr. Martnez is a specialist in radiology and medical imaging. She works in investigation and research at the University Foundation of Health Sciences (Fundacin Universitaria de Ciencias de la Salud - Fucs in Spanish). Her lines of research are Cardiothoracic and vascular intervention.
- Dr. Jos Hernndez: Dr. Hernndez is an neurologist expert in cognitive reserve PhD in cognitive functions.

Questions Reviewed:

The 4th group declared interest in the question of whether or not there are differences between the volume of the frontal lobe of the kangaroo subjects compared to non-kangaroo subjects. An approach without Braviz would include comparing the volume, particularly in the frontal lobe, in relation to the group using ANOVA.

Working Session:

For the resolution of the question using BRAVIZ the group created samples. Then proceeded to use the visualization tool (Surfaces) to find the name of the dataset variable that corresponded to the volume. After finding the variable the group used a linear model and analyzed the correlation between memory and volume of the frontal lobe. The group spent the end of the session trying to find variables related to *social cognition* and memory.

Difficulties and results:

At the end of the session the group was able to answer the initial question using several features, including cortical surface visualizations and ANOVA. Also, during the session, new questions were raised but could not be answered due to lack of information in the dataset. During the session the group had difficulties finding the variables needed and had some interaction problems while trying to use the same sample and selected variables in multiple statistical applications.

F. Group 5

Group Members:

- Dr. Juan Pablo Garca: Dr. Garca is a specialist in radiology and diagnostic imaging. He works in investigation and research at the University Foundation of Health Sciences (Fundacin Universitaria de Ciencias de la Salud - Fucs in Spanish). His lines of research are Cardiothoracic and vascular intervention, and neurology. His recent research projects include prognostic value of initial chest radiograph in pediatric patients with lower respiratory infection admitted to the emergency unit at Hospital San Jos.

- Angela Gmez: Angela Gmez is a doctoral student of psychology at University of Los Andes with an emphasis in neuroscience. Her research fields are consolidation of sleep-dependent memory , learning and language. She had previous experience with data analysis tools like SPSS and Stata. She has also used visualizations for support of her research with a software from the university called PSGLab used for analyzing sleep signals.

Questions Reviewed:

The 5th group reviewed the question *Do Kangaroo Infants show differences in their cognitive skills, and attachment quality with respect to children born on term?*. A typical workflow for answering the question included the use of ANOVA or a t-test to compare mean performance on all cognitive and behavioral tests between the groups.

Working Session:

In the working session the group searched for the values needed filtering by name and making several questions about available variables. After they found them they used the ANOVA tool. At the end of the analysis they found interesting reviewing the tools for detecting outliers.

Difficulties and results:

This group found enough time to finish the statistical analysis of the variables and draw conclusions on the question. Although they explored the brain visualization for the outliers the time of the session was not enough to ask new question regarding what they observed in this visualizations or to complement their initial conclusions.

G. Iteration 2 (Groups 3, 4, 5)

For the general discussion the groups found that the main difficulty with the tools comes at the time of selecting and finding variables. Some suggested that the tool could provide a model to interact with the variables in terms of the question and the concepts behind them.

The main interest was on how to interact with data sets compiled and organized by other researchers, how to import or export data and how to use and understand subsamples created with the tool. In particular, more metadata for the subsamples would be useful. Dr. Hernandez was interested in importing his own datasets to the tool to use in his research related to social cognition and memory.

Finally the participants suggested a adding video tutorials with a practical approach that could give them more time to explore the tool while adapting to their schedules.

H. Group 6

Group Members:

- Alejandro Osorio Forero: Alejandro Osorio is a graduate assistant at University of Los Andes at the Biomedical Engineering Department. His main lines of research are Cognitive Neuroscience and Signal Analysis. His research included a work arround Sleep Neurophysiology and a Classification of sleep stages from polysomnographic recordings using feature selection methods and unsupervised clustering.
- Karen Eliana Corredor : Karen Corredor is a psychology doctoral student at University of Los Andes. She works in behavioral neuroscience and in the development of a model of environmental enrichment for recovery of adverse life experiences at early life using an animal model (Wistar rat). When interviewed about her use of analysis software and data visualization she noted that she performs analysis of behavioral tests using specialized software. She also performs histological analysis of the brains of experimental subjects and uses image processing to detect quantitative and qualitative differences in tissue and brain structures.

Questions Reviewed:

This group of the last session was interested in determining if there is a relationship between the intervention and attention deficit or hyperactivity; and if there is one, which anatomical related to this relationship. The group mentioned that a traditional approach for them, without using BRAVIZ, would be taking a sample of both populations and looking for the relationship between these variables. For the quantitative part, they would use a software analysis of MRI. They did not specify which tools they would use for both parts of the question.

Working Session:

In the working session Karen tried to identify socioeconomic variables that had a correlation with household characteristics, but, could not find one with enough significance to continue analyzing brain volume data. The group then decided to change the question to one related to the relationship between depression and attention deficit in the kangaroo group.

The group could easily create the filters for the samples of kangaroo children with good quality brain images and after this they performed an ANOVA analysis with attention deficit as the outcome and depression as regressor. They could not complete the test because the values appeared to be aliased. Unfortunately, time was not enough to draw final conclusions.

Difficulties and results:

This group did not interact with the brain visualizations that corresponded to the second part of the question as they were not able to complete the statistical analysis. Even though the main question was not answered the group performed several statistical analyses to search for variables related to household. They found the interaction with the dataset difficult especially at the time of finding the needed variables.

I. Group 7

Group Members:

- Santiago Zarate: Santiago Zarate is a doctoral student of psychology at the University of los Andes. His research areas are Neuroscience and Behavior. When asked for his use of data analysis tools he declared that he uses software like Stata and SPSS, and programming languages like Matlab and R. In his current work he analyses FMRI and other similar measures,

and so he was particularly interested in the use of the software.

- Angie Paola Correa : Angie Correa is a doctoral student of psychology at the University of los Andes. She works at the Neuroscience and Behavior Laboratory at the University of los Andes. In the laboratory they work with data records collected from electroencephalographic signal, behavioral data of rats, and physiological measures such as cell count (histology). They use a registration software for recording behavioral data and for analysis they use STATA, SPSS, Matlab and R.

Questions Reviewed:

The question reviewed was about whether or not the level of intelligence of children can be explained by their levels of depression and acute stress reactivity. For an approach without BRAVIZ the group decided to calculate the Pearson correlation using variables such as the scores of depression and responses in the fear conditioning task, using any statistical software (SPSS or STATA).

Working Session:

During the practical session the group selected the variables for the ANOVA analysis from the dataset using a table of variables provided at the beginning of the session. After obtaining the results, they proceeded with the exploration of graphical and statistical results to determine the significance of the effects. At this point, BRAVIZ proved to be an useful communication tool since they could show each other graphical evidence of their conclusions.

After finishing with the analysis of the first question they spent the last part of the session in the exploration of other BRAVIZ applications such as multi-subject fMRI. Findings about particular questions or data were not obtained, as this was only an exploration of these applications.

Difficulties and results:

At the end the group had enough time left after answering the question for exploring additional features of BRAVIZ .The main difficulty for the group was finding the variables because of the large amount. Also they noted that the applications offer different paths to the same goal and experience is needed to know which path is better.

J. Iteration 3 (Groups 6, 7)

At this session it was remarked that BRAVIZ tools have an advantage over traditional methods for analyzing data with programming languages. Braviz does not allow mistakes like trying to relate variables that can not be related. Participants also remarked that even though they had little experience with the tool the navigation was intuitive.

They also commented that the interface has the potential to be used in studies with animal models as the database can be fed with data from these investigations. They also suggested different video tutorials for teaching the tool, taking different practical approaches depending on the users speciality.

Finally, the interaction with the data was difficult and the group suggested a feature that allowed the creation of a *data workspace* that allowed filtering only variables interesting for the current session so the user would not have to deal with all the variables of a large dataset all the time.

V. Discussion and Future Work

For the future work arround BRAVIZ the test revealed several insights on the questions asked at the beginning of this report:

Is the tool successful in stimulating insights?

The success of BRAVIZ in stimulating insight was evaluated in terms of the capability of the users to generate answers to the questions asked, supported by the evidence of visualizations and data analysis tools. Because of the users inexperience with the software and data a significant amount of time was invested in questions around data available and the tool itself. Still, most of the participants were able to formulate questions, answer them, and formulate new ones.

The most commonly used tools were the sample creator, ANOVA, linear model and correlations. The used brain visualizations were the subject overview, for outliers of the ANOVA test, the Explore fMRI application, and one group tried to create a volume using the ROI Builder.

How does the tool support the identification of data anomalies before analyzing it?

Some users declared that the tool presented an advantage over the use of a programming language as the data analysis tools would let them check the variables first and avoid analyzing variables that *are not susceptible of being analyzed*. Additionally, Dr. Charpak declared that visualizations usually let them know which data is correct and which one is not.

Still, at the working sessions some users tried to use empty variables for statistical analysis, not looking at the plots of the data that were shown when they selected the variables. Also some users noted that they would like to see included measures of reliability of the brain data, for example for the volumes obtained by Freesurfer and imported to the tool.

Does the background of the users affect the performance in the use of the tool?

Although the statistical analysis modules were the most used by all users, there were different approximations to their use by the different specialities of experts invited to the test. Starting by the scope of the questions: psychologist were focused around the HOME, depression or anxiety test results while the radiologist and neurologists focused on the influence of the kangaroo project in specific volumes of the brain.

As a consequence of the scope of the questions few of the psychology students used the brain visualization tools for supporting their answers, even though some declared that for the scope of their investigation these features would be useful and showed interest in following the training to use BRAVIZ as a support for their investigation.

How does the tool help the process of hypothesis generation?

During the process of answering the questions, there was no a clear indication for users to to make explicit hypotheses

about the data. This was expected to be observed along the process in the context of the discussion of the two users. Most of the hypotheses were stated at the moment of interpreting the outputs of the statistical tests and deciding if there was enough support for them. While discussing the answers, the groups used BRAVIZ as a tool for communicating their point of views and describing their insights visually.

How does the tool help the communication of hypothesis and insight?

As stated above, BRAVIZ visualizations were an effective vehicle for the communication of evidence of the answers in a direct and intuitive way, by both, expositors of the idea and audience. The visualizations aided the creation of new questions and directed observations on the data as the users could share their interests on certain variables showing the plots of the data to their coworkers.

In addition to the questions asked at the beginning of the test the obtained feedback could be summarized in the following:

The main difficulty at the moment of facing the questions was translating them into variables. Users stated that the selection of the variables was the most difficult process as searching and filtering variables by keywords, descriptions or categorization of the variables was not supported. If users wished to interact with information sources where they are not aware of the whole structure of the dataset it is very difficult to find specific information as there is no way of obtaining an overview of the data structure.

Compared to other tools used by the participants BRAVIZ offers the advantage of easily and intuitively providing visualizations for statistical analysis without the need of programming or detailed customization. The users found the integration between statistical analysis tools and brain visualization tools very useful specially for the analysis of outliers.

VI. Conclusion

Conclusions on the methodology will focus on 4 points: difficulties, developments introduced, advantages and future improvements.

The main difficulties faced at the execution of the test were:

- The coordination of schedule with the experts had to be careful since they all had few available spaces, especially physicians.
- The dataset was the main point of difficulty for the methodology. Because users were not able to work with their own data, the process of understanding the new data set consumed a significant amount of time. The feedback of the session would sometimes be oriented to the data structure, organization and disponibility and not to the tool itself.
- When the training session and practical session were scheduled on different days some users were not able to come to both. This affected the evaluation process as they did not have as much knowledge of the tool as other users.
- The users declared that the time scheduled was not enough to acquire the knowledge necessary to operate the tool. Nevertheless they were able to use the basic applications and interact with them during the practical session.

Between session iterations changes were introduced to the methodology:

- The initial data presentation was complemented in the last session with an spreadsheet containing metadata of the KMC dataset. This document was useful for improving the search of variables and understanding their structure and notation.
- In the second iteration the training and work sessions were integrated to accommodate to the schedules of the participants. This proved to be an advantage for expert users who do not have much time disponibility.
- Different methods of documentation were tried along the evaluation. Recording voices in the individual computers and sharing virtual log documents with the users proved to be the most organized and practical method. Video recording the use of the tool and the discussion inside the groups was also an useful source of information.

Advantages:

- The main advantage of using a data set not known to the users was the ability to observe how the users interacted with data organized in a way not familiar to them as sometimes happens in research projects. This point helped understanding the challenges at the moment of interacting with new data, in contrast to other methods where the users are comfortable with the data.
- The methodology proved to be useful for our particular case where not many users were familiar with the tool. It was able to provide enough insight on the use of the tool without the need of extensive training or existing use cases.
- In contrast to quantitative methods where large amount of data is collected from several users this method is useful when the objective public is limited.
- Organizing the groups in pairs helped to review the process of communicating using the tool and still preserving the diversity of insights provided by the different questions asked by each group.
- The inclusion of different disciplines allowed us to evaluate the performance of the tool with different scopes of knowledge and research interests.

Future Improvements:

- For future training sessions the process could be complemented with video tutorials oriented to different disciplines that help users gain more experience with the tool.
- The interaction with the data has to be improved in order to reduce the feedback focused on data and not

the tool itself. An alternative is using data known by the users or providing another tool for exploring the data.

- Different disciplines can be included in the test and mixed in the groups to observe the interaction of a multidisciplinary team . This would also allow to evaluate more complicated communication interactions.

REFERENCES

[1] . Shneiderman y C. Plaisant, Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies, en *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, 2006, pp. 17.

[2] . Perer y B. Shneiderman, Integrating statistics and visualization: case studies of gaining clarity during exploratory data analysis, en *Proceedings of the SIGCHI conference on Human Factors in computing systems*, 2008, pp. 265274.

[3] . Saraiya, C. North, y K. Duca, An evaluation of microarray visualization tools for biological insight, en *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on*, 2004, pp. 18.

[4] . Lam, E. Bertini, P. Isenberg, C. Plaisant, y S. Carpendale, Seven guiding scenarios for information visualization evaluation, 2011.