PUC

How signification and communication systems influence the interpretation of statistical data by users with specific information needs

Taissa Abdalla Filgueiras de Sousa
Simone Diniz Junqueira Barbosa

Departamento de Informática

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO

RUA MARQUÊS DE SÃO VICENTE, 225 - CEP 22451-900

RIO DE JANEIRO - BRASIL

How signification and communication systems influence the interpretation of statistical data by users with specific information questions

Taissa Abdalla Filgueiras de Sousa

tsousa@inf.puc-rio.br

Simone Diniz Junqueira Barbosa

simone@inf.puc-rio.br

Abstract. Charts can represent and communicate quantitative information more efficiently than tables. However, their interpretation requires knowledge of graphic systems, which nowadays can be made more efficient with the aid of interactive computer systems. Based on Semiotic Engineering, this paper reports a study on how communication and signification systems represented by Web visualization tools influence the interpretation of charts. We investigate the users' abduction and sense-making process through the triangulation of two qualitative methods: Communicability Evaluation Method (CEM) and Think Aloud with co-participation. The first method allowed us to evaluate failures in the reception of the designer-to-user message, revealing features of the computational signs that hinder metacommunication. The second one allowed us to investigate the users' abductive processes during charts creation and to identify problems in understanding and generating chart using the investigated tools.

Keywords: Information visualization, semiotic engineering, signification, communication, abduction, sense making, communicability evaluation method.

1 Introduction

Pinker [10] affirms that charts are an effective means of communication, because they take advantage of human perception and cognition mechanisms. Shah e Carpenter [15] also describe that, in numerical data presentation, charts are used because humans are able to recognize visual patterns. Aligned to them, Cleveland [4] claims that quantitative patterns and relationships between data are quickly revealed by charts because of the power of our visual-cognitive system to realize geometric patterns.

Taking advantage of the characteristics of human perception, however, some resources can be used in graphic language to confuse the reader. Tufte [18] explains the most common problems: the absence or neglect of scale; the occultation of the initial value of the scale, which should always be zero (otherwise cause a disproportion between the compared values); and the comparison of an entire element with part of it (*e.g.*, compare values of the previous (whole) year values with current (ongoing) year values).

This type of communication requires specific knowledge from the "reader." According to Goldenberg [9], to interpret charts correctly, students need mathematical knowledge besides perceptual experience. Goldenberg [9], Clement [4] and Gomes [10] claim that chart interpretation requires knowledge of the graphic system, which is difficult because it includes rules that are not easily learned by readers. Our social problems increase the difficulty: high rates of functional illiteracy [13], lack of schooling on statistics in high school [3], and the lack of questioning by the reader, consuming information without careful analysis and becoming easily manipulated by the media [3].

Computational systems that allow users to build and interact with charts can influence data interpretation and learning of the graphic system. There are several systems designed to aid in the interpretation and to improve the user experience. To motivate the user interest for statistical data, informal visual representations are used by several of them, for instance: Manyeyes, GapMinder e Worldmapper. Others, such as Statplanet and MCLikertVis, bring together two or more visual representations. This research aims to investigate how interactive computational systems can make more efficient the interpretation of statistical data by users who want to answer a specific question. In our studies, we have focused on an educational theme, and from this theme we chose a question to be answered by our user, supported by some visualization tools available on the Web. Based on education research by Soares [16], the chosen question was: How have the cognitive abilities of youngsters increased in the last 10 years in Brazil and how is Brazil positioned in comparison to other countries? We found these data in the following Web tools: Statplanet, SIDRA and Statistical Series.

Grounded by Semiotic Engineering, our research question was to investigate: How do signification and communication systems influence the interpretation of statistical data by users with specific information questions?

The Semiotic Engineering concepts we chose to explore in this research were: signification and communication systems, abduction and communicability evaluation method (CEM). From these concepts, only CEM was developed specifically within semiotic engineering, for evaluating how users interact with computational systems; the other ones were inherited from Semiotics. To help triangulate the results, we have also selected the think aloud technique.

This research is organized as follows. Section 2 relates visualization tools with signification and communication systems. Section 3 explains the abductive process and relates it to our theme. Section 4 describes user experiences in both methods of evaluation with users (CEM and think aloud). Section 5 shows the results of the triangulation of the two methods and, finally, section 6 discusses about the users' abductive processes as they were perceived during evaluation.

2 Visualization tools and communication and signification systems

"Signification systems are the result of culturally conventionalized sign functions. In other words, whenever cultures produce a convention that correlates a certain set of expressive elements to a certain set of contents, a signification system is in place. Signification systems constitute the codes available for communication processes. These processes, however, are not constrained to expressions that belong to a signification system, but communicators may choose to use innovative or latent sign functions allowed by signals of the expressive plane that have the potential to participate in sign functions." [6, p. 72].

In a communication process, signification systems are ways to represent (expression) intent significations (content) to achieve a set of goals (intention) [1].

Thus, we can understand that the graphic system, represented by Cartesians charts, maps, cartograms or others, constitute signification systems. They are defined by mathematical, statistical, and geographical conventions that allow us to interpret information. In other words, they are codes used in the communication.

In other hands, visualization tools are interactive computational systems that use specified representations of the graphic system. Interaction with these tools allows the creation of new signs, which are supported by sign functions of the graphic system. Each one of these tools also has its own language constituted by its own signification system, an artificial language made by signs of the interface and processed signs through the interaction between user and system. These signs communicate the designers' message to the user about why and how he or she can or should use the system and with which effects.

Visualization tools are, therefore, communication and signification systems at two different levels: by the set of generated visualizations and by the set of signs produced by the interaction designer. The communication process involves, in the representation(expression): the signification (content), that is, the set of available data; and the goals (intentions), that is, the intention to answer questions about several data simultaneously and to stimulate learning of the graphic system. The new signs, created by the interaction, allow the designer to influence the user's understanding of the graphic system and the user's understanding of the message. We explain this further in the following sections.

The communication and signification systems chosen were: Statplanet, SIDRA, and Statistical Series. They are all Web-based tools. Statplanet provides information at the level of countries and it is used to create info charts and cartograms. It also aims to be an educational software. SIDRA and Statistical Series are tools from "Instituto Brasileiro de Geografia e Estatística – IBGE," the Brazilian Statistical and Geography Institute, and provide Brazilian data at two geographic levels: one to to retrieve aggregated data, and another one to retrieve and show statistical information of periodic surveys.

Based on these three tools, we investigate how they influence the interpretation of users with specific information questions and which signs hinder the user's abduction process.

3 Abduction and sense making

There are three types of inference: deduction, induction, and abduction. In deductive inferences, what is inferred is necessarily true if the premises from where the inference arises are true. This means that true premises guarantee truth of the conclusion [17].

According to Santaella [14], abduction is the reasoning that leads to the adoption of a hypothesis, and induction is the test of that hypothesis. Also according to her, in 1901, Peirce defined abduction as the acceptance or creation of a smaller premise as a hypothetic solution to a syllogism having the bigger premise known and the conclusion found out as a fact. [14, p. 92]. This concept was soon extended to the view that: "it consists of the examination of a set of facts that allows these facts to suggest a theory" [14, p. 92].

Furthermore, Peirce says: "The explanation must be a proposition such as it would lead to the prediction of the observed facts, either as necessary consequences, or at least as very probable under certain circumstances. A hypothesis, then, must be adopted as plausible in itself and making facts also plausible. This step of adopting a hypothesis as suggested by the facts is what I call abduction." [14, p. 93].

To de Souza [6, p. 43], "the interesting fact, for computer science, is that abduction reasoning does not depend on the existence of known and stable primitives or axioms. The elements that function as primitives and axioms are hypothesized, produced on demand, and tested for suitability against the evidence collected by the human interpreted (or reasoner)." In our research, it is interesting to separate two levels of abductive reasoning. At the first level, we find the information visualization techniques that graphically represent data of a certain domain. At this level, the user might generate hypotheses related to the best graphic representation of data. At the second level, we find the artificial language constituted by the system's own communication signs. At this level, the user needs to learn to use the system through an acquisition process that involves the recognition of symbolic patterns. Both levels involves corrections of the user's hypothesis on how to use the system based on the understanding of the language that is acquired and refined during user-system interaction.

Bertin [2] defines efficiency of representation by the following proposition: "If, in order to obtain a correct and complete answer to a given question, all other things being equal, one construction requires a shorter observation time than another construction, we can say that it is more efficient for this question" [2, p.139]. Bertin continues: "the most efficient constructions are those in which any question, whatever its type and level, can be answered in a single instant of perception, that is, in a single image." [2, p. 146].

Interactive computational systems can aid the user's abductive reasoning in the search for efficient representations and understanding of information more easily.

However, the user must understand the system language. If the hypothesis formulated by users converges to efficient graphical constructions and corrects analysis, we can consider the communication between the designer of the system and the user to be efficient.

Our goal is separate the two abductive processes in a way to allow the analysis of these two levels of interpretation: the visualization tool and graphic system.

4 User Evaluation

In this research we aim to identify user difficulties and misinterpretations while using visualization tools, so we can propose new interface solutions that help converge their interpretations. We used an endogen triangulation of two qualitative methods of user evaluation: the Communicability Evaluation Method (CEM) and Think Aloud Protocol with co-participation (involving two users simultaneously). Using the first method we aim to evaluate failures in the reception of the message by the user, revealing features of the computational signs that hinder meta-communication. Using the second one we aim to investigate the user's abductive processes during charts creation and to identify problems in understanding and generating charts through the use of tools, by motivating the verbalization of their thoughts and the conversation between users.

4.1 Preparation

The preparation stage was the same for both methods. We defined the user profile as people with ability in reading charts. Thus, we chose students who had just been enrolled in college or who were in its first period of a STEM course. The ability to interpret charts was an important requirement, so we screened participants for that ability. They must also have never used the evaluated tools. To find out about their familiarity with map-based tools, a desirable characteristic, we asked about their ability in Google Earth an Google maps. They all had ability with charts, had not used the evaluated tools before, and were used to Google interface systems. When asked about the use of software for creating charts, some of them declared having used Microsoft® Excel.

Table 1 describes the selected participants.

Table 1: Users profile

Method	Evaluation Code	User Profile
pilot test	A00	17-year old student in the second period of Information Systems. Uses computers to develop systems, to surf on the internet, to write texts and to create tables and charts with Microsoft® Office.
CEM	A01	18-year old student in the second period of Information Systems. Uses computers to search in google, to communicate via social networks, to write and to answer emails, and to work on academic coursework.
	A02	180year old student in the second period of Information Systems. Uses computers to conduct research.
	A03	20-year old student in the third period of Computer Engineering. Uses computers to study and for fun. Among the most frequently used softwares are Photoshop, After Effects and Firefox. The user has used Excel to build charts.

Think aloud with co-participation	D01	D01-1) 18-year old student in the first period of Computer Science. Uses computers to search on the internet, to write and to answer emails, and to watch videos on youtube. Among the most used softwares is Visual Studio. D01-2) 19-year old student in the first period of Computer Science. Uses computers as entertainment and to develop systems. Among the most used softwares are Visual Studio and Office.
	D02	D02-1) 18-year old student in the first period of Computer Science. Uses computer to play videogames, to develop systems and to download movies, songs, and TV series. Among the most used softwares are Visual Studio, 3D, and iTunes. D02-2) 20-year old student in the first semester of Computer Science. Uses computers to play videogames, to develop systems and to search on the internet. Among the most used softwares are Visual Studio, Google, and Office.
	D03	D03-1) Computer Engineer, 37 years old. Uses computers to develop systems using the C programming language. Also uses Excel. D03-2) Computer Engineer, 24 years old. Uses computers to play videogames and to develop systems.

After the pilot test, we made few adjustments in the scenario to emphasize the most important tasks for our analysis. The final scenario used in CEM and Think Aloud evaluation was:

Paul is a researcher for the government. He analyzes statistical data to provide technical and institutional support to the governmental actions for formulation and reformulation of public policies and to Brazilian development programs. He basically works with educational data. Today, he needs to participate in several meetings, so he left his trainee, John, in his place to find the data to make a research. Paul sent the following e-mail to John in order to let him understand about the subject and ask for some instructions:

Dear John,

The Programme for International Student Assessment (PISA) is a worldwide study by the Organisation for Economic Co-operation and Development (OECD) repeated every three years to conduct a large-scale educational assessment among its member countries. The sample of PISA is composed of 15 and 16-year old students that attend at least the 7th grade of a formal teaching institution. The first edition of PISA was conducted in 2000 and the last one in 2009. This period coincides with the moment in which the Brazilian education systems have been adopting cycles to reduce failure and dropout rates in schools. Despite the expansion of the universe the PISA sample, there were changes in the yield of Brazilian students in successive editions of PISA.
We need to verify how the cognitive abilities of young have increased in the last 10 years in Brazil and how Brazil is positioned in comparison to other countries.

In order to answer this question, I need you to create charts to:

- Observe the variation of score in 31 countries that participated in the survey for two periods: 2003 to 2009 for trends, and 2006 to 2009 to compare recent results. Identify the rise of the national average and the countries that had the most positive development in this period.
- Show the evolution of the Brazilian scores in PISA.
- Show that there was an increase of students able to take the PISA exam in 2000, 2003, 2006 and 2009, or that there was an increase in schooling among 15 and 16-year old

teenagers during in this period.

You can use some online visualization tools for retrieving data and building the charts: Statplanet, SIDRA, and Statistical Series. Remember that the last two systems are from IBGE and you can get only data from Brazil. The IBGE surveys where you can find about schooling are PNAD and Censo 2000.

With this scenario, we ask them to obey the following instructions: Generate the described charts in the scenario, export and save the result in a Word document. If possible, save your result in the system itself. If you do not find exactly what is being asked, you can use data that allow you to reach a similar conclusion. If necessary, you can explain your results in the document.

To perform the task, the 31 countries [12] mentioned in the scenario were listed and attached to the given task, as follows:

1. Australia
2. Austria
3. Belgium
4. Brazil
5. Canada
6. Czech Republic
7. Denmark
8. Finland
9. France
10. Germany
11. Greece
12. Hungary
13. Iceland
14. Ireland
15. Italy
16. Japan
17. Latvia
18. Liechtenstein
19. Luxembourg
20. Mexico
21. Netherlands
22. New Zealand
23. Norway
24. Poland
25. Portugal
26. Republic of Korea
27. Russian Federation
28. Spain
29. Sweden
30. Switzerland
31. United States

We used Techsmith® Camtasia software to make the recordings.

The first two tasks can be performed using Statplanet. The third task can be performed using SIDRA either or Statistical Series. The following links were opened in different browser tabs:

- Statplanet - http://www.sacmeq.org/statplanet/StatPlanet.html

- SIDRA (PNAD) - http://www.sidra.ibge.gov.br/pnad/pnadpb.asp?o=3&i=P

- Statistical Series - http://seriesestatisticas.ibge.gov.br/

Figure 1 shows the scatterplot, as designed by Soares and Nascimento [16], whose research formed the basis for our scenario. The scatterplot is a type of point chart [5]. It is also an exploratory method useful for providing a first look at a given bivariate to see how it is distributed across the plane, for example, to see clusters of points and discrepancies [8]. The figure shows the variation of the average of the disciplines in PISA 2000 and 2009. The diagonal splits the chart into two equal parts. In the area above the line we find the countries with positive growth. Below the line we find the countries with negative variation. On the line we find the countries without variation. Thus, the growth of the score in each country corresponds to the length of the line drawn vertically from the point to the diagonal. In this chart, we can see that Brazil had a more positive change than almost all other countries, except Luxembourg.
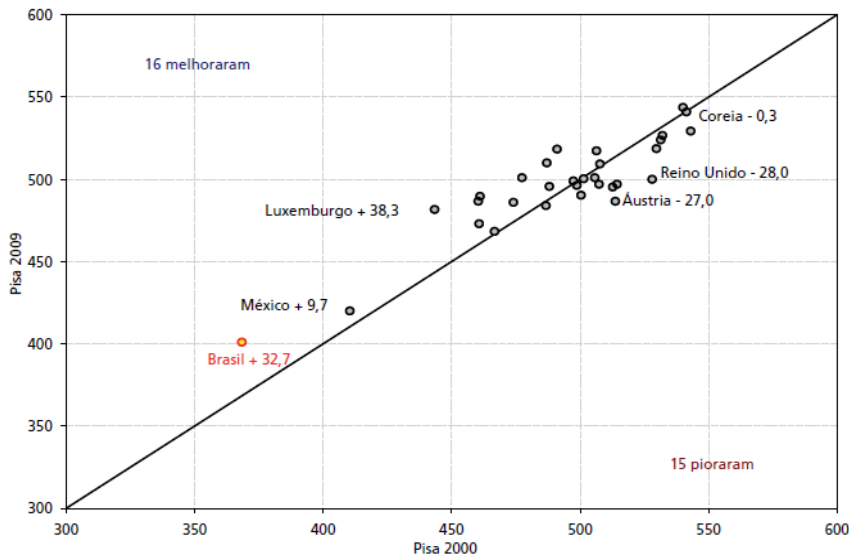
Figure 1: Scatterplotby Soares e Nascimento [16].

The first two items of the task can be performed using Statplanet by choosing one of the subjects (reading, math or science), since the system does not have the average of the three disciplines. The third task can be performed using SIDRA or Statistical Series. Figures 2, 3, 4, 5, and 6 show the expected charts to respond our question using these systems.

Figure 2: Scatterplot designed in Statplanet with a comparison between the reading scores of PISA 2003 and 2009. To compare 2006 and 2009, the chart would be similar, the only change would be in the horizontal axis.
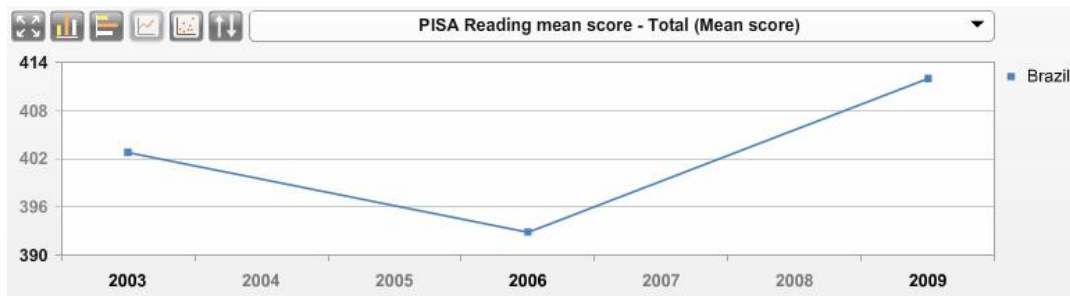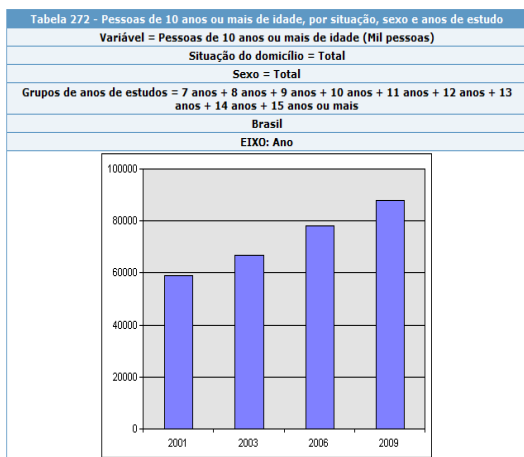
Figure 3: Chart for the second task

Figure 4: First result for the third task, using SIDRA.
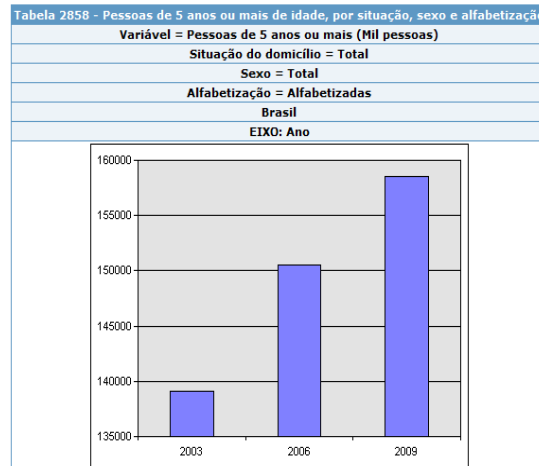

Figure 5: Second result for the third task, using SIDRA.
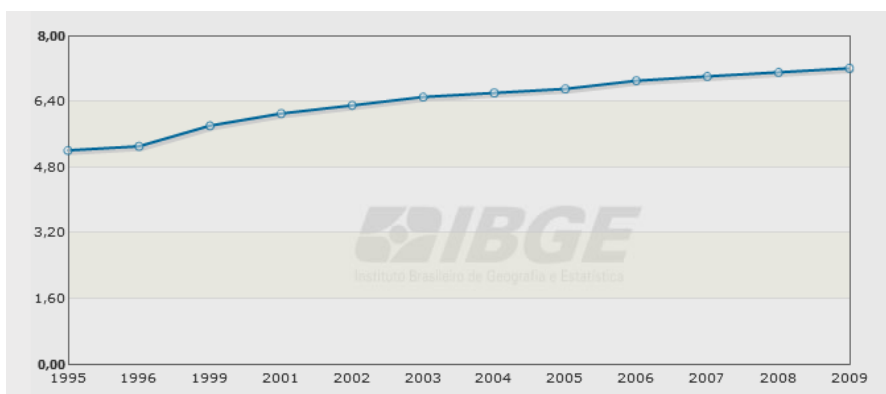

Figure 6: Chart for task 3 using Statistical Series.

The goal of CEM is to evaluate the quality of the message as received by users. First, we conduct a brief inspection of the user interface signs with the reconstruction the designer's message and the identification of inconsistencies, ambiguities, and contradictions. Then, we observe and collect data while user interacts with the system. Next, we tag and interpret the communicative breakdowns in the video, elaborate the semiotic profile and, finally, consolidate and report the results.

4.1.1 Inspection
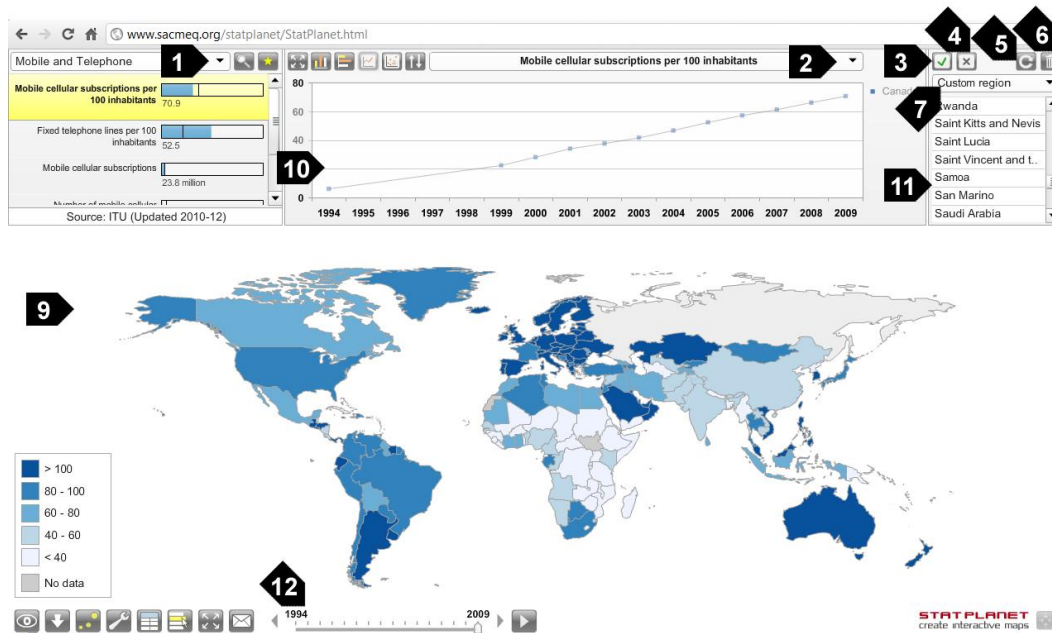
Tasks 1 and 2: Statplanet

Figure 7: Statplanet.

To check whether the user may have problems understanding the tool, let us recall the first task of our scenario. We can perform it by finding the math scores theme (Education > Quality education – OECD > Pupil achievement reading, math, science > PISA mathematics mean score - Total); selecting the requested countries (through customization, deletion or addition of countries); locating the scatter plot; selecting year 2003 in one axis and year 2009 in the other and, finally, exporting the chart. For the more recent comparison, we need only to change 2003 to 2006 in the correspondent axis.

Table 2: Inspection of inconsistencies, ambiguities and contradictions in the first task.

Problems	Where?
Inconsistencies (Figure 7: Statplanet.)	The selection widgets (elements 1 and 2) have different contents. In element 1, the selection widget shows one level of information below of the content of element 2. For instance: In 1, you can only find "Pupil achievement reading, math, science" but, in 2, the system allows the selection of the last level of information, the scores. Furthermore, by default, the system loads the first item of the selected menu in the map and in the chart, which can confuse the user.
	Excluding countries: the label of element 5 is "Remove selected countries" and the label of element 4 "is" Deselect all countries." For the line chart, both operators work in the same way as the countries of the graph are selected. This distinction only makes sense for the bar charts because, by default, they start with all the countries that have the selected data.
	There is no way to undo a deletion. The only way is through the icon "Refresh selection list" (5), which includes all countries again, deleted or not.
	In our scenario, the user could try to find the data by clicking "PISA (OECD) Participating countries", which hides the charts of the tool, keeping only the map with marks in the participating countries.

	There is no feedback when a country on the map is selected. The feedback occurs only when the mouse is positioned over the country. However, the selected countries cannot be observed on the map after the click.
Ambiguity (Figure 7: Statplanet.)	Region customization: the instructions are not clear (elements 3,4,5,6, and 7). They allow the user to have different interpretations about the order and steps to customize a region.
Inconsistency (Figure 7: Statplanet.)	There is no legend to inform the meaning of the colors of the points in chart.
	The choice of years is made in different ways for each axis: using a combo on the X axis (13), but a slider on the Y axis (12).
	The data visualization is difficult because Statplanet traces a line that is the linear regression of the values (14), and not a diagonal, as seen in Figure 1. If the user does not realize this, as Brazil is below the line, he could reach an incorrect interpretation that Brazil had a negative change.

We have not found problems in performing the second task. The chart can be created as shown in Figure 3.

Task 3: Sidra and Statistical Series

These two tools can be used to construct the chart to show the increasing schooling between young Brazilians of 15 to 16 years old in the period of interest.
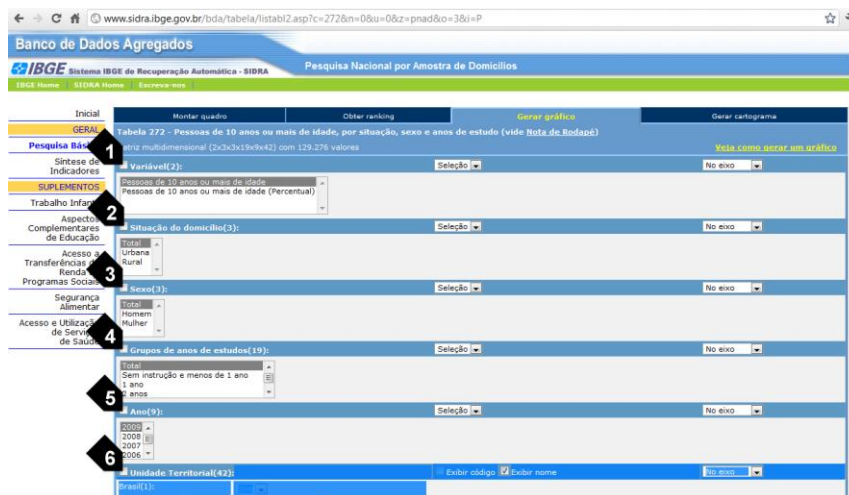

Figure 8: Interface for chart construction in SIDRA. Elements 1 to 6 represent the selectable dimensions.

In SIDRA, without changing the default chart type of the tool, we can create the charts shown in Figures 4a and 4b to perform the third task. There are two options for themes: People 10 years old and older by state, sex and number of years of study (Figure 4a); Persons 5 years old or older by state, gender and literacy (Figure 4b).

Table 3: Inspection of inconsistencies, ambiguities, and contradictions in the third task with SIDRA.

Problems SIDRA	Where?

Inconsistencies (Figure 8)	Nothing is reported about the maximum number of selectable dimensions to create a chart (You can select all dimensions together - 1, 2, 3, 4 , 5, and 6). This is different from the explicit message of the designer, which allows only two dimensions for the chart. It becomes clear only through the error "Select only two dimensions for the chart."
	The chart type is chosen after the selection of the variables at the end of the screen, *i.e.*, the user does not know which chart will be generated when choosing the variables.
Ambiguity (Figure 4 and Figure 5)	There are problems in the constructed charts: the scale of the graph in Figure 5 does not start at zero. It can give a false impression that the increase of literate people was much higher than what is shown in the left chart (Figure 4). The designer should make it difficult to generate charts that can confuse the reader.

Figure 9: Statistical Series. Element 1 corresponds to the menu, and element 2, to the search.

The chart shown in Figure 6 could be generated through Statistical Series (Figure 9) by selecting the education theme and then "Average of schooling years of the population of 10 years and older by gender".

Table 4: Inspection of inconsistencies, ambiguities, and contradictions in the third task with Statistical Series.

Problems in Statistical Series	Where?
Inconsistencies (Figure 6)	The chart in Figure 6 can cause erroneous interpretations because it does not have a regular spacing in the horizontal axis. For example, the data from 1996 to 1999 appear to have been

	growing faster than the other years because there is a two-year gap in the data collection. Another problem is the note with the explanation about the reason why there are not data in certain years. It is not associated with the chart, but placed at the bottom of page.

4.1.2 Collection

Only two users (A01 and A02) finished the three tasks. A03 finished only the first two tasks. However, all of them had difficulty in analyzing the obtained charts and answering the questions of the task with confidence. All users used Excel at several times during the evaluation to support the chart construction. Another relevant point was that no one completed the third task using SIDRA. A01 and A02 gave up quickly, switching to Statistical Series. A03 insisted in using SIDRA, consulted the available help but also gave up.

About the results of task 1, A01 opted for the column chart (Figure 10) from the beginning, A02 passed by the column chart and returned to the line chart (Figure 11) and A03 browsed all available charts in the tool and also opted for the line chart (Figure 11). The region customization was tried by A02, but failed.
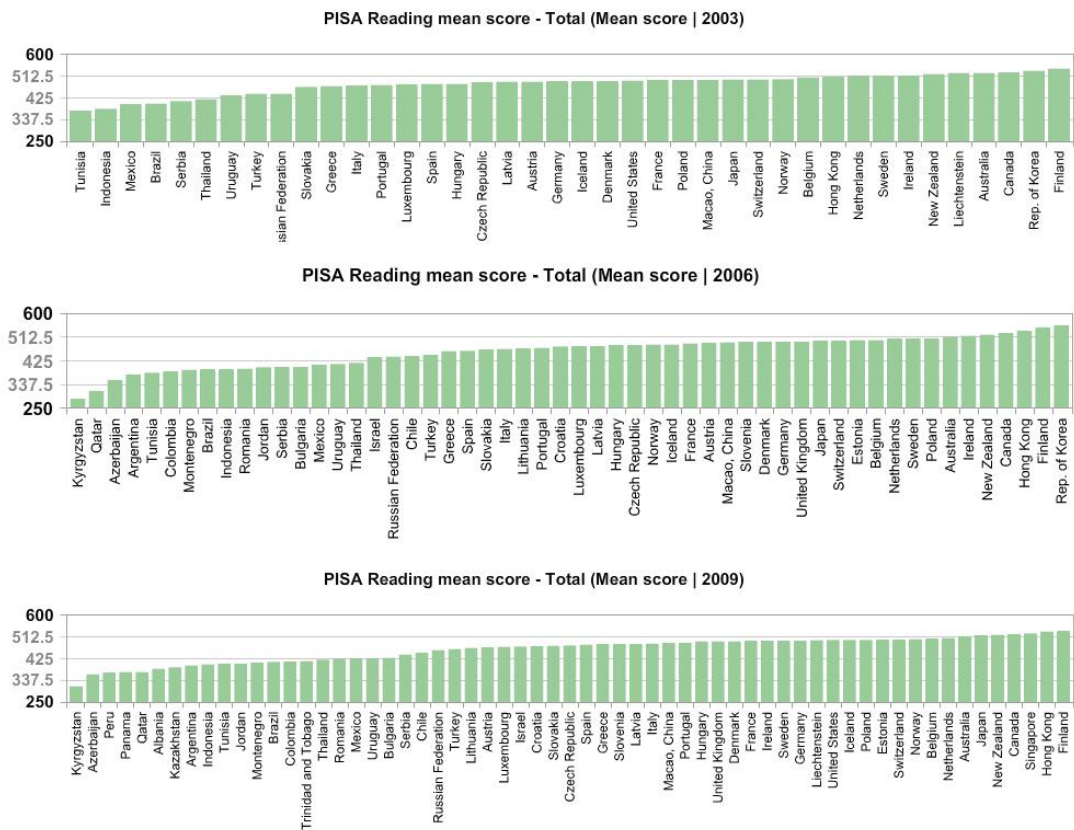

Figure 10: Chart 1 by A01 - The result were 3 bar charts of the years 2003, 2006, and 2009
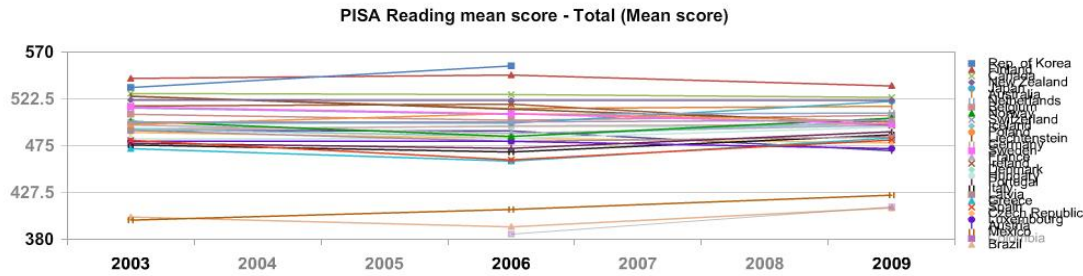
PISA Reading mean score - Total (Mean score)

Figure 11: Chart 1 by A02 and A03.

About the results of task 2, A01 generated the column chart using Excel with exported values from Statplanet (Figure 12). A02 selected three column charts as shown in Figure 13. A03 performed the task successfully and easily obtaining the expected result (Figure 3).
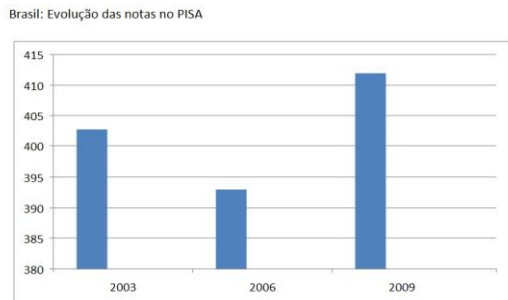

Brasil: Evolução das notas no PISA

PISA Reading mean score - Total (Mean score | 2003)

PISA Reading mean score - Total (Mean score | 2006)

PISA Reading mean score - Total (Mean score | 2009)

Figure 12: Chart 2 by A01 - The result was made in Excel. The tool does not allow trend graph in columns.
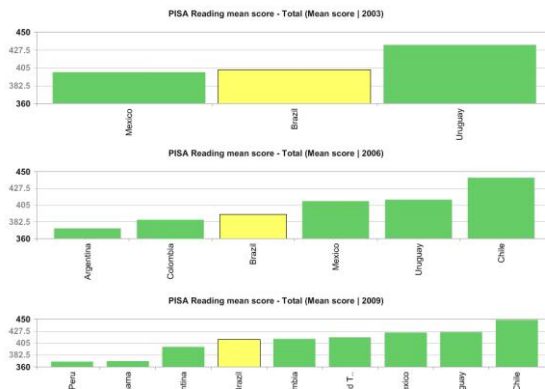
Figure 13: Chart 2 by A02 - A02 could not put the 3 columns together in the same chart with the tool.

No user has created charts for the third task. A01 and A02 found the data in Statistical Series but failed to export the chart and decided to export the data table. The result presented were Excel tables.

4.1.3 Tagging and Interpretation

CEM can identify communication breakdowns, "moments of interaction in which the user demonstrates not having understood the designer's metacommunication, or moments when the user finds it difficult to express their communicative intent communication" [1, p. 346]. These disruptions are tagged by communicability expressions, which are questions or expressions in natural language with precise technical meaning. A summary with the identified breakdowns is presented in Table 5. This table shows the types of breakdowns by user and by frequency: the more asterisks, the more often was the disruption. The shaded columns represent the most severe disruptions. Note that, from the 13 tags defined by CEM, only 11 were identified. The labels "Why doesn't it?" and "Where am I?" did not occur in the tests.

Table 5: Tagging

	Statplanet	Statistical series	SIDRA

		A01	A02	A03	A01	A02	A03	A01	A02	A03
Temporary failures 1. User's sense making is temporarily halted	1. Where is?	***	*	*	*** **	***	*	*	*	*
	2. What happened?		***	*						
	3. What now?			***		*	***			***
Temporary failures 2. User realizes her intended interaction is wrong	4. Oops!	*	*							
	5. I can't do it this way.	***		***						
Temporary failures 3. User seeks to clarify the designer's deputy's intended meaning	6. What's this?	*** *** *	*** *** *	*** *** ***		*				
	7. Help!	*								*
Partial failures	8. I can do otherwise.	***	***		*	*				
	9. Thanks, but no, thanks.				*	*				
Complete failures	10. Looks fine to me.	*	*	*						
	11. I can't do it.							*	*	

The tags of complete failure were found only in SIDRA and Statplanet: **"I can't do it"** occurred in SIDRA for users A01 and A02 when they have chosen to quit the tool and try the other one. **"Looks fine to me"** was a tag found in the evaluation of all users of Statplanet. They have completed the chart and considered the task done, but we classified in this way because there were better ways to present the data with Statplanet that were not seen by them. Although two of them (A01 and A03) have declared, in the post evaluation questionnaire, that they were satisfied with the generated charts, they all said in the interview that they would have done it differently if the tool so allowed. The level of satisfaction declared was because they believed they did the best they could do with the tool.

Partial failures were more frequent. **"Thanks, but no, thanks"** was found only once in Statistical Series. A01 and A02 opted to the search by keywords to find the theme instead of the desinger's preferred solution, the classification by theme. **"I can do otherwise"** was found in the chart export functionality in Statplanet as well as in Statistical Series for all users. In Statistical Series, as it did not allow exporting charts, both users (A01 e A02) exported the data table to generate charts in Excel. A02 also removed the columns that were not necessary. In Statplanet, A01 exported three column charts (one for each year) for the first task and then tried to put them all together in Excel. The same tag happened in Statplanet when A01 tried to customize a region and then tried another solution. A02 chose the list to select countries. This was not the designer's preferred solution. A02 also declared not to be satisfied with this task and answered in the interview that he would like to have created a single chart with three columns to show

the scores of Brasil in PISA 2003, 2009, and data variation. A02 also would like to remove the Latin American countries.

Temporary communication failures are less serious problems. They are classified into three types: (i) those that interrupt the user semiosis, (ii) those in which the user realizes that his communicative act was not successful and (iii) when the user tries to understand the system's speech act.

Among the tags of semiosis interruption (i), **"Where is?"** was the most common in all tools and for all users, happening many times during the evaluation. We could see it, for instance, at the moments of seeking for themes in the three systems, of changing the year, of improving the legibility of generated charts and of inserting countries in chart.

"What happened?" occurred in the countries selection by A02 and A03. A02 also got this tag while trying to customize regions. **"What now?"** happened to A03. A03 switched several times between the three systems for a few minutes, showing that he did not know what he should do.

"Oops!" was found with less often than **"I can't do it this way."** These tags are used when the user realizes that his communicative act was not successful (ii). "Oops!" occurred mainly in Statplanet, when A02 unintentionally deleted all selected countries, and also when A01 exported a map instead of the chart he wanted to export. "I can't do it this way" occurred when A01 attempted to find the same theme in a different year in the menu of themes, and when A03 tried to include countries in the chart through direct manipulation from a data table and in other ways that the tool does not allow.

The third type of temporary failures occurs when the user wants to understand the communication process (iii). We had "What's this?" and "Help!" Usually, "What's this?" happens after "Where is?" or "What now?" tags. We have seen successive **"What's this?"** tags in Statplanet, when users were looking for something. In these moments, users have repeatedly positioned the mouse over each sign in order to see the corresponding explanation and to find the desired function. **"Help!"** happened in the evaluation of A03, when he used SIDRA and looked for explanations in links of help available on the system interface. It also happened to A01 when he accepted to receive help from the evaluator to find the issue. We considered it as an aid because we had offered it before and it was refused.

4.1.4 Semiotic Profile

After the tag interpretation, we elaborated the semiotic profile of the tools. "The semiotic profiling stage finalizes the whole process of communicability evaluation. In it an in-depth characterization of the designer-to-user metacommunication is achieved. The designer's message can be spelled out by the evaluator, who will be able to assume the first person in discourse and say the following: Who do I think are (or will be) the users of the product of my design? (...) What have I learned about these users' wants and needs? (...) Which do I think are these users' preferences with respect to their wants and needs, and why? (...) What system have I therefore designed for these users, and how can or should they use it? (...) What is my design vision? (...)" [6, p.146-7]. We found that there are differences between the metamessage issued by the designer and the meta-message received by the user. Below we describe the emitted metamessage of each tool rebuilt in the step of the inspection and marked in bold the points that were not well received by users.

Statplanet's Designer-to-user Meta-message

You are a professional from statistics or economics, and may even be a student interested in visualizing statistical data at the level of countries. You can also be a person who has no experience in these matters, but who is interested in using charts. You should have knowledge of tools that use maps such as Google Maps.

You want to visualize such data from charts and cartograms because you can easily spot trends and relationships between data. You also need to generate tables, export data, charts and cartograms. You may even want to change some settings, such as changing chart colors, replacing names by acronym. **You prefer to view the data through the map** and use the chart to complement your research and see how they are related. **You also prefer to select countries through the map because you easily identifiy a country in the map**. **The alphabetical list of countries is a secondary option** that you may want to use to find smaller countries or others you do not know where they are localized.

Thus, you should start by selecting a theme to create a cartogram and at least one country to visualize data in the chart. You can animate the generated cartogram to see how the data have evolved over the time, or even to view the temporal information in a line chart. You can change the chart to bar or column to visualize the comparison between countries. **You can create charts to correlate two sets of data and to identify relationships between them**. You may want to select only a certain area of the map, so we included some predetermined regions and **we enable you to create your own set of countries**. Every time you position the mouse over a chart element or a country, all information related to the element will be highlighted, so you will be able to easily analyze it.

SIDRA's Designer-to-user Meta-message

You are a professional of statistics or economics. **You know the IBGE surveys and know the data that are be available in each survey.** You want to retrieve and manipulate aggregated multidimensional data. You may want visualize them as tables, rankings, charts or cartograms to develop your research. You need to handle aggregate data as needed. **Thus, to generate a chart, you must select two dimensions to compose the chart axes and with as many classes you want. Classes can be displayed, filtered or summed.** Other dimensions may only have one selected class. You can also choose between **different chart options** the best suited to for the interpretation of the data that you want to show. We offer 21 different chart types, including bar chart, column chart, time series, polar chart, among others. Each table has up to 6 dimensions.

Statistical Series's Designer-to-User Meta-message

You are a student or even a researcher interested in Brazilian socio-economic and demographic data. You want to recover these data from a given survey, theme or geographic region. You may want to see that data through a direct query. You prefer to consult them mainly in the form of tables and charts, but you can also want to generate cartograms. You want to see the series evolution, so we built a tool that allows you to view series charts. You do not need to construct the series because it is ready to use. **You only have to select the topic, data source or geographical level,** and then select few parameters such as data type (percentage or absolute), geographic area, unity, and territorial categories.

4.2 Think Aloud with co-participation

The think-aloud instructions are usually very short, referring to a procedure which is presumed to be familiar to users. For instance: "Try to think aloud, I guess you often do so when you are alone and working on a problem" [7]. In order to make this protocol as natural as possible to the user, we decided to form pairs (coparticipation [19]). The

pairs were formed by the volunteers and were usually constituted by friends or colleagues. The only instruction given was to talk while executing the tasks.

4.2.1 Data Collection and interpretation

We observed that the conversation became more agile at times but, in general, it has not helped to perform the task. However, our test achieved its goal, as the conversation among the participants revealed abductive processes of each pair, especially from the leader, who usually talked more to explain his movements to the other participant. The secondary participant, however, had often had his abduction process stopped for not having direct access to input devices and had to accept the explanations of his colleague. In this review, none of the pairs has finished the task, leaving the third task pending.

Using Statplanet to perform task 1 and 2, Table 6 shows where the doubt has motivated them to search for a new explanation in their abductive processes and Table 7 shows disruptions in interpretation of the graphic system.

Table 6: Communicative breakdowns in Statplanet

Breakdowns	D01	D02	D03	Users' utterances
Hard time finding the theme	*		*	*"There is no PISA!"* D03-1 trying to find the theme.
Hard time finding the chart after selecting the wrong theme "Participating countries of Pisa", which hides the chart			*	*"But the chart is enabled"* D03-1 checking the settings to find an explanation.
Breakdown in the selection of countries	*	*	*	"How do we select 31 countries?" D02-1 asking D02-2 about task 1. *"There is about 50 countries! Let's count it!"* D02-2 noting that they had selected more countries than requested. *"He could allow adding countries. It's easier to add 31 than to delete the rest".* D03-2 talking about the selection of countries.
Breakdown in customizing regions.	*	*	*	*"It killed everything!"* D03-1 when he unexpectedly lost the selected countries. *"I want to create a custom chart but it does not.* D02-1 trying to customize regions. *"Try the custom again!"* D02-2 after other attempts in the selection. *"It's locked!"* D03-1 clicking on the customize button without feedback.
Breakdown in exporting the chart and use of print screen button	*			*"How are we going to take the chart to Word? Is there an option to save a chart? - Print screen"* D01-1 and D01-2 trying to export the chart. *"Let's look for something to save it"* D02-2 to D02-1

				looking for an export button.

.

Table 7: Communicative breakdowns in the graphic system

Breakdowns	D01	D02	D03	Users' utterances
Doubt in choosing the Chart	*	*	*	*"A chart like that will be huge, it is better to change to bars"* D01-1 about line chart. *"Should we try bars or lines? - Bars!"* D01-1 and D01-2 about the final decision. *"There is a trend here, dude! Ok?"* D03-1 to D03-2 about switching from bar chart to line chart.
Conceptual problems and misinterpretation	*			D01 often reached miscalculations about the score variation.

Figure 14 and Figure 15 represent results from D01 e D02 to the first task. D03 created a chart similar to Figure 11. The second task was performed by all pairs as expected (Figure 3).
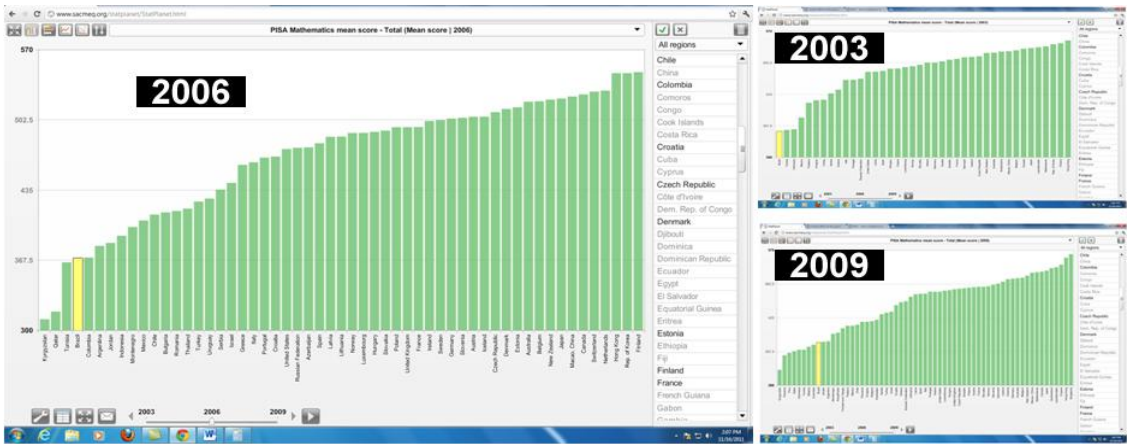

Figure 14: Chart 1 by D01 – They created one column chart for each year (2003, 2006, and 2009) and selected Brazil in each one.
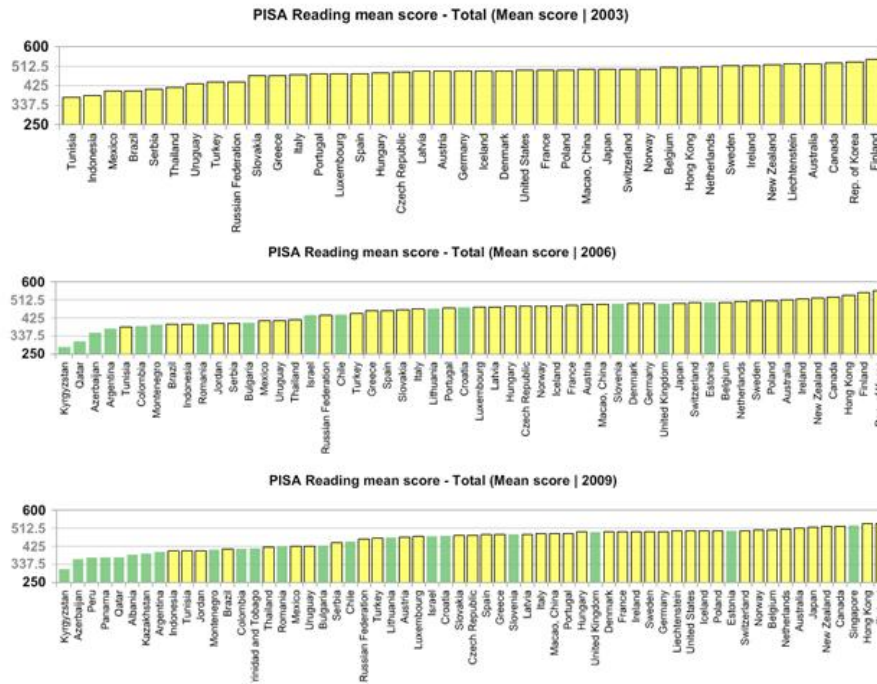
Figure 15: Chart 1 by D02 – They created one column charts to each year (2003, 2006, and 2009) and selected (yellow) the 31 countries of the scenario.

5 Discussion

The two methods we used allowed us to find different results: CEM revealed several breakdowns in all tools. We observed, through the complete breakdowns that led users to dropout from SIDRA, that those users have failed to formulate any explanation about the grammar of the tool's language in their abductive process. Users who resorted to Excel (Statplanet and Statistical Series) also did not understand this artificial language.

However, we identify some signs that hindered the chart generation by users. Among these signs, we find the countries selection through the map and the customization of regions (Statplanet), where recurrent breakdowns. Also, the preferred type of selection by the users were secondary options (Statplanet and Statistical Series). In these signs, the metamessage received by the user diverged from the one issued by the designer. After some time of using the tool, they reached an explanation about its features, but the charts generated still diverged from what they expected. This suggests that breakdowns in the communication with the system caused breakdowns in the understanding of the charts grammar.

This method, however, did not allow us to understand how and why the user chose those visualizations, whether the user would know how to interpret those charts, or if he has perceived mistakes in some constructions before the post-test interview.

The think-aloud technique complemented our research because it enabled us to understand more about the users' abductive processes as well as their mental models, and to identify problems in chart interpretation. During the execution of task 1, observing the users' abductive processes, we conclude that signs represented by computational charts were determinant for the choice of representation by the user.

As the grammar of the Statplanet language was different for each chart type (bars and columns are filled with all countries by default while line charts start empty, awaiting user selection, the scatter plots needed a selection of another theme for the y axis), some users, moved by an opportunistic behavior, found the convenience of the column charts and chose this option, generating three charts. This solution could be efficient if the selected countries in each graph were the same. However, other users, who had in their mental model that trends are constructed with line charts, kept their models, which resulted in inefficient visualizations. The fact that the line chart is the default chart in its interface was also a factor for the user to choose this one.

We have also seen that the difficulty in selecting and customizing regions in Statplanet led to mistakes. Users have inferred that the enabled countries were the countries of the scenario, which was not true, since every year new countries join the list. In SIDRA, we observed spontaneous comments that strengthened the communication breakdowns and users' difficulties identified in previous method.

In the post-test interview, with the presentation of two charts: scatterplot (Figure 1) and clustered column chart (Figure 16), we saw that users were not as familiar with the first one, and so did not arrive at this result, which would be the most efficient representation available in the tool. The second one would be an option as efficient as the first, but it was not available in the tool. When we asked what chart (including the one they had done) would be the best to represent what was requested in task 1, they answered that the clustered column chart was more familiar and that they would like to have this option.
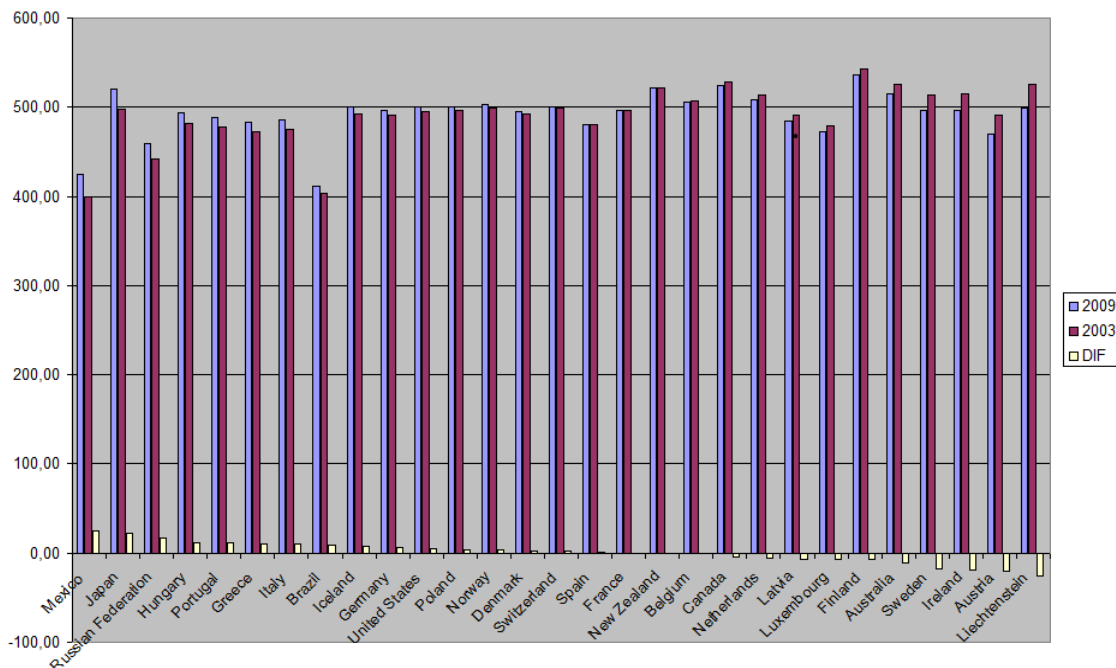

Figure 16: Clustered Column Chart. The first two columns are the scores of the requested years and the third column shows the difference between them.

In the following section, we present the collected results in the post-test questionnaires (Table 8) and show our analysis based on results obtained and on users' answers in interview.

Table 8: Answers to post-test questionnaire. NA = Not Applicable

	CEM			Think aloud		
	A01	A02	A03	D01	D02	D03
Task 1: (Easy) 1 - 5 (Difficult)	3	4	2	4/4	2/3	3/3
Task 2: (Easy) 1 - 5 (Difficult)	1	3	1	4/3	1/3	1/3
Task 3: (Easy) 1 - 5 (Difficult)	1	3	4	NA/NA	5/5	3/4
Satisfaction (Y/N)	Y	N	Y	Y/N	Y/Y	Y/N
Did you want more help? (Y/N)	Y	N	Y	Y/Y	N/Y	Y/Y

O item 1 ficou classificado como intermediário, tendendo mais a fácil para alguns usuários e mais difícil para outros. Isso porque alguns tiveram dificuldade em manusear o Statplanet nos primeiros minutos de uso. "Achei a ferramenta Statplanet inicialmente confusa de usar; é preciso ir clicando e observando os resultados, mas depois fica mais tranquilo" (A02). "Depois que aprende fica fácil e ela é boa. Depois de alguns minutos analisando ela você começa usar ela muito bem" (D02-1). Outros declararam ter sentido falta de manuais: "Deveria haver algum recurso que pudesse auxiliar a pessoa durante o primeiro uso da ferramenta" (A1). "Faltou um guia rápido de como usar" (A03). D02-2 declarou que os textos explicativos eram muito curtos. "Faltou um pouco de observações. Por exemplo: Remover selecionados. A gente não sabia o que era selecionado".

The first task was classified as intermediate, easier for some users and harder for others. This happened because some of them had difficulty to handle Statplanet in the first few minutes of use. "I thought the tool Statplanet initially confusing to use; we need to go clicking and looking at the results, but then it becomes easier" (A02). After learning it is easy and it is good. After a few minutes analyzing it you start to use it very well" (D02-1). Others said they wanted tutorials: "It should have some resource to support the person during the first use of the tool" (A1). "Missed a guide on how to use it" (A03). D02-2 stated that explanations were very short. "Missed some observations. For example: Remove Selected. We did not know what was selected."

The second task was considered easy and average by most users. Users who classified the task between average and difficult affirmed they had already had more familiarity with the tool in this task because of item 1, but even then did not classify the task as easy.

In task 3, users who used the SIDRA rated the item with a difficulty level much higher than the participants who used Statistical Series. Note, however, only one user scored 1 (easy) to the task. This one was A01, who ignored that the obtained series did not contain the years 2000 and 2009, and even spawned the requested chart. Even A02, who also used Statistical Series, considered the task as average because he had trouble finding the theme and still had to use Excel.

In the issues about satisfaction with the generated charts, the result was balanced: half of the users were satisfied and the other half were not. A02 and D03-2 were not satisfied and they justified their answer by the lack of readability of trend charts. How-

ever, A03 and D03-1 were satisfied because they created a trend chart as their understanding about the requested in the task.

Users who made the composition of three charts were satisfied with the obtained graphics, except for D01-1, who did not think that the data were correct. He said: "It might have some way to generate the average of the 3 disciplines." Actually, all presented charts did not allow a correct analysis of the comparison because the number of participating countries increased over the years of the survey. However, D02 could be satisfied. They marked all participating countries in yellow. Visualizing Brazil would be possible to make a correct judgment of its rise.

SIDRA was considered the most difficult tool. A03, D02 and D03 tried to use it, but failed to complete the task, even though they had access to manuals. A03 exclaimed - "The SIDRA is impossible!" D02 gave up saying: "I cannot understand it! Go to Google!"

6 Conclusion

We conclude that the methods used were complementary. CEM allowed finding communication breakdowns between designer and user and differences between the issued and received metamessages, as well as identifying signs that hindered the computational generation of charts. The think-aloud method reinforced some points found by CEM and still allowed us to identify the signs that influenced users in choosing the constructed charts.

Regarding the tools, we found that, when the user had more visualization options, he had more breakdowns in the reception of the message. In this research we found severe problems of understanding the designer's message through SIDRA, the tool that allows more sophisticated data manipulation, where we can perform operations as sum, average and filters, and build more than 20 different types of graphical representations. Statplanet demonstrated some flexibility, making some tasks possible such as filtering countries, visualization with time line and correlation between two indicators, but for functions related to chart constructions it is quite limited. By the other hand, Statistical Series was considered easy, but it is limited to temporal series data.

Through the analysis of these systems of meaning, we conclude the language and grammar of visualization tools have great influence on the visualizations generated by users. If there are many breakdowns in the reception of the message by users, this is reflected on the graphical constructions and hence on the interpretation of the data.

Acknowledgments

References

[1] Barbosa, Simone D.J. and Silva, Bruno S. (2010) Interação Humano-Computador. Campus/Elsevier.

[2] Bertin, Jacques. Semiology of Graphics: Diagrams, Networks, Maps. 1918. ESRI first edition 2011, USA

[3] Cazorla, Irene. A relação entre a habilidade viso-pictórica e o domínio de conceitos estatísticos na leitura de gráficos. Tese (Doutorado em Educação). UniCamp. Campinas, 2002.

[4] Clement, J. "Misconceptions in Graphing". Proceeding 9nd Annual Meeting of the International Group for the Psychology of Mathematics Education, 1985. p. 369-375.

[5] Cleveland, William S. Visualizing Data. AT&T Bell Laboratories, Murray Hill, New Jersey (1993).

[6] de Souza, Clarisse S. The semiotic engineering of human-computer interaction. Cambridge: Mass. The MIT Press. 2005. Chapters 1, 2, and 4.

[7] Ericsson, K. Anders; Simon, Herbert A. Protocol Analysis. MIT, 1993.

[8] Few, Stephen. Show me the numbers. Designing tables and graphs to enlighten. Analytics Press, Oakland, California. 2004

[9] Goldenberg, E. Paul. (1988) Mathematics, metaphors, and human factors: Mathematical, technical, and pedagogical challenges in the educational use of graphical representation of functions. The Journal of Mathematical Behavior, Vol 7(2), 135-173.

[10] Gomes Ferreira, V. G. Exploring Mathematical Functions Through Dynamic Microworlds. 353 f. Tese (Doutorado em Educação). Institute Education, Universidade de Londres. Londres, 1997.

[11] Pinker, Steven. (1990). A theory of graph comprehension. In R. Freedle (Ed.), Artificial intelligence and the future testing (pp. 73-126). Hillsdale, NJ: Erlbaum.

[12] PISA 2000 – Relatório Nacional. Brasília, 2001 p. 19. Available at: http://darwin.futuro.usp.br/site/doprofessor/Relat%F3rio_PISA2000.pdf [last access on 20/Nov/2011]

[13] Portal do Professor. Available at: http://portaldoprofessor.mec.gov.br/fichaTecnicaAula.html?aula=23488 [last access on 10/Jan/2012]

[14] Santaella, Lucia. O método anticartesiano de C. S. Peirce. Editora UNESP, 2004.

[15] Shah, Priti and Carpenter, Patricia A. (1995). Conceptual limitations in comprehending line charts. Journal of Experimental Psychology: General, 124(1), 43-61.

[16] Soares, Sergei S. D.; Nascimento, Paulo A. Evolução do Desempenho Cognitivo do Brasil de 2000 a 2009 face aos demais países. Instituto de Pesquisa Econômica Aplicada, 2011.

[17] Stanford Encyclopedia of Phylosophy, Abduction. Available at: http://plato.stanford.edu/entries/peirce/#dia [last access on 23/Oct/2011]

[18] Tufte, Edward R. The Visual Display of Quantitative Information. 2001

[19] Wilson, C. (1998). Usability techniques: Pros and cons of co-participation in usability studies. Usability interface, 4(4).