

An Infrastructure to Support the Visual Analysis of Data from the Brazilian Public Health System

Carlos H. A. do Nascimento^{1*}, Glauco de Figueiredo Carneiro^{1*}, Ed Santana M. Costa¹, Ethan V. Munson²

¹ Universidade Salvador (UNIFACS), Salvador, BA, Brazil.

² University of Wisconsin-Milwaukee, Milwaukee, WI, USA.

* Corresponding authors. Email: carloshelano@unifacs.br, glauco.carneiro@unifacs.br

Manuscript submitted April 1, 2017; accepted August 12, 2017.

doi: 10.17706/jsw.12.8.612-630

Abstract: Understanding how Brazilian health units have been reimbursed due to services provided on behalf of the Brazilian Public Unified Health System (SUS) is not a simple task. It requires the analysis of a myriad of data and hampers the identification of flaws as well as improvement opportunities regarding the budget application in the context of health public services. In this paper, we proposed an infrastructure to support the visual analysis of data from SUS related to services provided by licensed health units. To illustrate the use of proposed infrastructure, we conducted an example of use that analyzes the performance of licensed health units from the Brazilian Northeast region in a specific time interval.

Key words: Brazilian public unified health system, health information systems, information visualization.

1. Introduction

Public health plays an important role for citizens and has drawn attention of both policymakers and the public at large. Concerns about gaps in both the availability and quality of public health services have hit the headlines in response to both new and persistent health risks [1]. For this reason, decision makers and practitioners must be careful about the allocation, management, and administration of public health resources [2].

The effectiveness of health care expenditure is a worldwide concern and this is not an exception in Brazil [3]. Brazil's citizens expect the public health system to make appropriate use of out-of-pocket health expenses. In 1988, within the new Constitution, the Brazilian Unified Health System (SUS) was created by the Brazilian Ministry of Health to offer free health care to the population, based on three pillars: universal coverage, integral health care and equity [3]. In 1991 the Information Systems Department of SUS, called DATASUS, was created to support the Ministry of Health in developing and maintaining software systems to process data in the health domain. Table 1 presents the software systems maintained by DATASUS. The Decentralized Hospital Information System (SIHD) registers health unit admissions. Data related to these admissions are provided monthly by the health units that provide services to citizens on behalf of SUS¹. The SUS health unit Information System (SIHSUS) supports the management of the medical billing invoices sent by health units from all parts of the country that provide services on behalf of SUS². This software system deals with invoices related to the cost of stays in health units and is the focus of analysis in this paper. The main reason for this choice is the considerable amount of data related to the topic and the

¹ <https://goo.gl/ObVSdg>.

² <https://goo.gl/2ecW4d>.

corresponding specificities. Another system, called TABWIN, supports health unit staff with data analysis. It allows the tabulation of dBase files (DBF) that constitute the main components of SUS information systems. TABWIN manipulates different types of information such as health unitization data, mortality, and population. Users can download data from the site of DATASUS.

Table 1. Systems Provided by the Brazilian Ministry of Health Related to Medical Billing Invoice

Goal	Application	Database	Local/ Network
Medical Billing Invoice	Decentralized health unit Information System (SIHD)	Firebird	Local
	SUS health unit Information System (SIHSUS)	ORACLE	Network
Validated Medical Billing Invoice	TABWIN	DBF/Text	Local
Supporting Parameters	National Registry of Health Institutions (CNES)	Firebird/Text	Local
	SUS Medical Procedures and Medicine System (SIGTAP)	Text	Local

We have found studies reported in the literature that analyze budget application in the context of health public services in different countries³. For example, the World Health Organization provides worldwide data related to different health themes. However, there is no specific report to present worldwide health-care expenditure by country⁴. To the best of our knowledge, in the Brazilian scenario, there is no software suite available to help citizens to keep track of SUS expenditures. The software systems provided by the Ministry of Health are not integrated with one another and the way data is available requires a thorough knowledge of the data set structure and organization. In one of these studies, the authors analyzed atypical occurrences in terms of availability and accessibility to public health services in Slovenia [4]. For this purpose, they applied data mining techniques and visualization resources to represent the results. The use of data mining techniques requires the availability of a significant amount of data as exemplified in the study reported in [5]. In that work, the authors described how data mining resources can support public health decision makers in the analysis of data regarding brain tumors. We have observed that despite the amount of data to be analyzed, there are not many studies describing the effectiveness of visualization resources in the analysis of data related to the public health sector. To fill this gap, this paper proposes an approach for supporting visual analysis of data from the Brazilian Public Health System.

This paper is organized as follows. Section 2 contextualizes data available from the Brazilian Public Unified Health System. Section 3 presents some of the advantages of information visualization as a tool for data analysis. Section 4 describes the proposed infrastructure to support the visual analysis of health data. Section 5 discusses the rationale behind the visual metaphors to support the analysis. Section 6 illustrates the use of the visual approach to analyze a subset of the Brazilian Health Data. Section 7 presents conclusions and future work.

2. Data Available from the Brazilian Public Unified Health System

The vast majority of Brazilian citizens depend on the public health system. For this reason, the analysis of health system data deserves special attention. To place the analysis in context, we created a process flow shown in Fig. 1. It was adapted from the Portuguese version of the Hospital Information System Guide from the Brazilian Unified Health System (SUS)⁵. The process flow describes the steps required to approve a

³ <https://goo.gl/BTdRsN> ; <https://goo.gl/d5R6rX> ; <https://goo.gl/U3I9ui>

⁴ <https://goo.gl/KA5YWx>.

⁵ <https://goo.gl/1FCkL5>.

medical billing invoice sent from the health units for approval by the Brazilian Ministry of Health. It begins when the health units' staff members access a local DATASUS System called SIHD (described in Table 1) daily to register data to generate the medical billing invoices. The next step is the validation of the invoices by this local system and the monthly generation of files containing data related to the aforementioned invoices. This text file is the input to a second level of validation performed by the State Health Department for the state where the health unit is located. In case the validation is not successful, the invoice is sent back to the health unit. Otherwise, the validated invoices are monthly sent to the Ministry of Health to be processed. The Ministry of Health then publishes the decision regarding the invoices approval or not on a web site⁶. This process flow involves different systems and different actors including staff in the health unit, the State Health Department and the Ministry of Health.

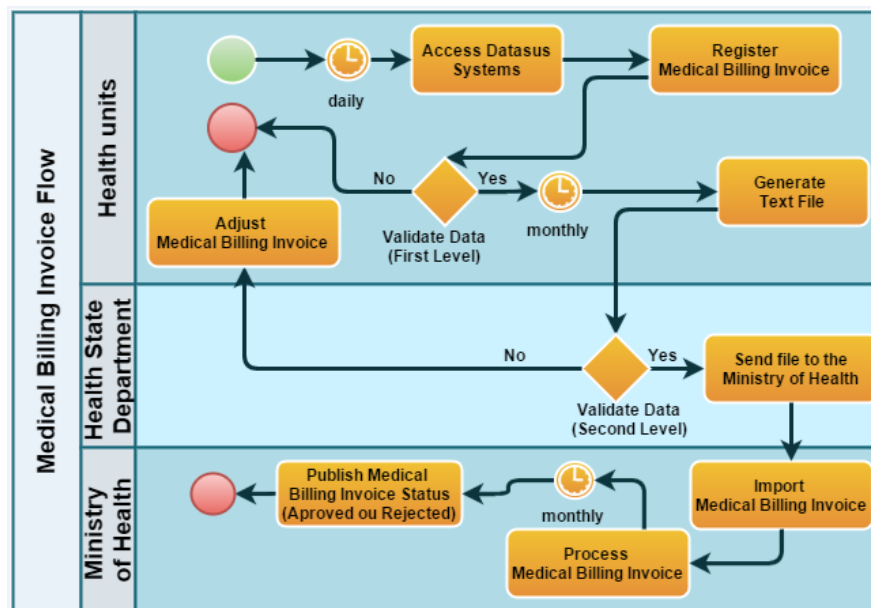


Fig. 1. The medical billing invoice process flow.

Among the challenges faced by health units in executing the steps described in Figure 1, reliable Internet service is not widely available. For this reason, some health unit staff cannot have instant access to data and applications provided by the Ministry of Health. In addition to the need of data availability, applications to support data analysis should be intuitive enough to meet user's needs. Unfortunately, this is not the case with the current data and applications provided by the Ministry of Health.

The Brazilian press has reported that the Brazilian Northeast region is among those regions that have suffered most from a shortage of public medical care provided by the Brazilian Public Unified Health System (SUS)⁷. Table 2 presents data on approved and rejected health unit accounts from the Brazilian Northeast region registered in January and February of 2014. This data set enables the identification of information such as the health units that have the highest percentages of approved and rejected accounts. The rejection situations are the result of inconsistencies in the procedure records sent to the Ministry of Health. After the analysis of Table 2, the Brazilian states with the highest number of approved accounts are Bahia (BA), Pernambuco (PE) and Ceara (CE) totaling almost three hundred thousand accounts. The respective monetary values for the corresponding approved accounts were R\$ 102, R\$ 107 and R\$ 83 Million of Brazilian Reais. In this period, despite having fewer approved accounts than the state of Bahia, the state of

⁶ <https://goo.gl/QpnD77> ; <https://goo.gl/X0OH0j>.

⁷ <https://goo.gl/rLmMj5> ; <https://goo.gl/1CZEGW>.

Pernambuco received more funds than any other Brazilian Northeast states.

The account rejection rate is just one of the indicators that can support analysis of Public Health expenditures in Brazil. Of course, this data set is not enough to show that the rejection rate variation is the result of the procedures provided in the health unit assistance. For example, surgical procedures are more expensive than other types of procedures. Besides that, not all health units are qualified to perform such procedures. This requires a more refined analysis to understand why these variations occur. The data presented in Table 2 were obtained after a sequence of manipulations of the files available at the DATASUS site ⁶. However, the data presented in this table is not enough for the analysis. For example, the table does not provide data to identify the health units that received the highest values for performed procedures among the Brazilian Northeast states. Moreover, the table also does not help identify the procedures that most contributed to these highest values. These are just a few examples to show the limitations of Table 2. For this reason, the proposed infrastructure presented in this paper supports appropriate operations to create an input data set that will be sufficient for a more informative analysis. Furthermore, this data set can be analyzed more effectively when supported by visualization tools, as discussed in the next section.

Table 2. A Sample of Approved and Rejected Accounts from the Brazilian Northeast Health Units

State	Month	Accounts Submitted	Corresp. Valued (Brazilian Millions R\$)	Accounts Approved	Accounts Approved (%)	Corresp. Value (Brazilian Millions R\$)	Accounts Rejected	Accounts Rejected (%)	Corresp. Value (Brazilian Millions R\$)
BA	201401	68659	59.61	61631	89.8	52.14	7028	10.2	7.47
	201402	66800	57.78	60114	90.0	49.95	6686	10.0	7.83
Subtotal		135459	117.39	121745	89.9	102.09	13714	10.1	15.30
PE	201401	43985	56.39	42567	96.8	53.56	1418	3.2	2.83
	201402	44085	57.10	42523	96.5	54.26	1562	3.5	2.85
Subtotal		88070	113.49	85090	96.6	107.82	2980	3.4	5.67
CE	201401	43920	51.84	39697	90.4	43.87	4223	9.6	7.97
	201402	41668	47.63	38101	91.4	40.06	3567	8.6	7.57
Subtotal		85588	99.47	77798	90.9	83.93	7790	9.1	15.54
MA	201401	35469	25.79	34002	95.9	24.64	1467	4.1	1.15
	201402	33608	24.10	31877	94.8	22.81	1731	5.2	1.29
Subtotal		69077	49.90	65879	95.4	47.45	3198	4.6	2.44
PI	201401	19433	14.23	18889	97.2	13.89	544	2.8	0.34
	201402	17250	13.05	16872	97.8	12.83	378	2.2	0.23
Subtotal		36683	27.29	35761	97.5	26.72	922	2.5	0.57
PB	201401	17798	18.70	16004	89.9	16.11	1794	10.1	2.58
	201402	16507	17.77	15007	90.9	15.52	1500	9.1	2.25
Subtotal		34305	36.46	31011	90.4	31.64	3294	9.6	4.83
AL	201401	14969	14.84	13963	93.3	13.61	1006	6.7	1.23
	201402	13994	14.19	12932	92.4	12.42	1062	7.6	1.77
Subtotal		28963	29.03	26895	92.9	26.03	2068	7.1	3.00
RN	201401	14675	16.42	12847	87.5	13.91	1828	12.5	2.51
	201402	14628	17.78	13014	89.0	15.15	1614	11.0	2.63
Subtotal		29303	34.21	25861	88.3	29.06	3442	11.7	5.15
SE	201401	8064	7.29	7413	91.9	6.48	651	8.1	0.81
	201402	7588	7.93	6964	91.8	6.95	624	8.2	0.99
Subtotal		15652	15.23	14377	91.9	13.42	1275	8.1	1.80
Total		523100	522.46	484417	92.6	468.16	38683	7.4	54.30

3. Information Visualization as a Tool to Support Data Analysis

Many researchers have pointed out the important role and the powerful means that visualization plays in

interactive data analysis and information exploration [6]-[8]. Humans have the natural ability to track and detect visual patterns and this ability can be exploited to improve the data analysis effectiveness. Visualization resources can support the analysis of data set structures in situations where the analysis requires more than seeing only a brief summary of data [9]. One of these situations is when exploring the data to identify possible patterns, both to confirm expected ones and find unexpected ones. Another possibility is when assessing the validity of a statistical model, to judge whether the model in fact fits the data [9]. Interactivity is a crucial resource provided by visualization environments to handle complexity. A single static view can show only one aspect of a data set. For some combinations of simple data sets and tasks, the user may only need to see a single visual encoding. In contrast, an interactively changing display supports many possible queries [9], [10]. Depending on the size of the data set, the limitations can be overcome through interaction allowing user actions change the view the way forward [9].

According to the Card reference model for information visualization [11] represented in Figure 2, the pipeline of a visualization infrastructure goes through a sequence of successive steps. The process starts with raw data from a repository (part A). The data then undergoes a set of transformations to be organized into data structures suitable for information exploration (part B). This step corresponds to the pre-processing and data transformations represented in the transition from part A to B. Next, the data structures are used to assemble visual data structures (part C). Those structures organize data properties and visual information properties in ways that facilitate the construction of visual metaphors [11]. This step defines the mapping from real attributes – which are derived from the data properties, in our case health data attributes – to visual attributes such as shapes, colours and positions on the computer canvas. This step corresponds to the visual mapping in the transition from part B to C. It is important to highlight that this step does not deal with rendering, but rather with building suitable data structures from which the views can be easily computed and rendered [12]. The last step is the view instantiation aimed at drawing the information on the computer screen to produce the views (part D). That is when a specific visual scene is rendered on the computer screen (part E). The model emphasizes that the visualization process is highly interactive [10], [12].

According to [9], information visualization usage and interaction is based on seven key categories listed as follows: Select (*mark something as interesting*); Explore (*show me something else*); Reconfigure (*show me a different arrangement*); Encode (*show me a different representation*); Abstract/Elaborate (*show me more or less detail*); Filter (*show me something conditionally*) and Connect (*show me related items*).

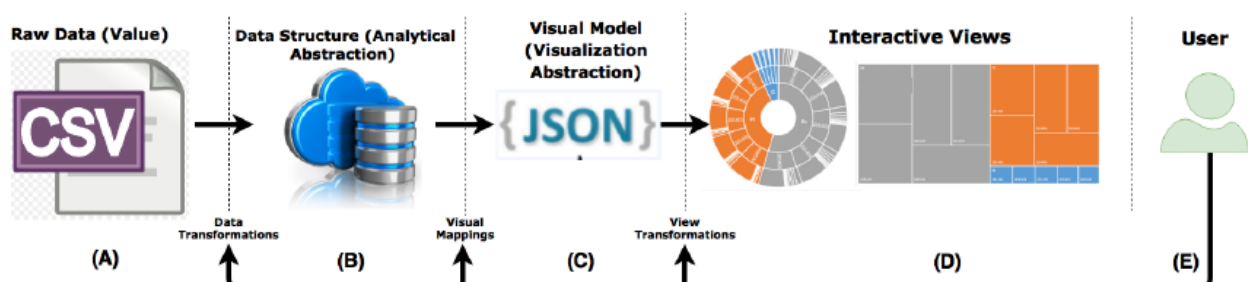


Fig. 2. An information visualization reference model adapted from (Card *et al.*, 1999).

We decided to tackle the challenge of analyzing the Brazilian public health data implementing the Card model for information visualization [11] and therefore providing users the opportunity to search for relevant information using the seven categories proposed by [10]. To this end, we proposed the infrastructure [10] as presented in Figure 3. We describe the infrastructure in the next section.

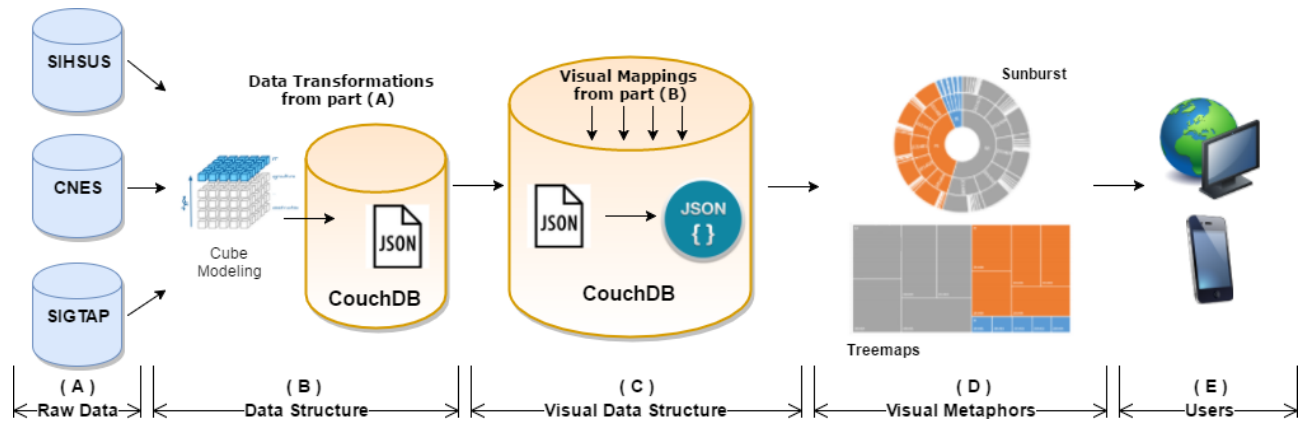


Fig. 3. A proposed infrastructure to support the visual analysis of the Brazilian health data.

4. An Infrastructure to Support the Visual Analysis of Health Data

This section describes the proposed approach to support the visual analysis of data related to health-care expenditures represented through visual metaphors. This infrastructure is shown in Figure 3.

The first part of the infrastructure deals with transformation from raw data to data structures stored in a database (Part A in Fig. 3). The data source in this case is the DATASUS site ⁸ that contains the files related to medical billing invoices created by SIHSUS. The raw data is available in compressed files and can be visualized in its original format using the TABWIN application. This same application enables users to export data in text format and to generate basic graphics. The transformation of the original data consists of converting data to approved and rejected medical billing invoices in JSON files where data is grouped by year, month and the Brazilian states in which the health unit was located.

Besides the aforementioned files, for a consistent analysis of services provided by health units, it is still necessary to download files that contain the unique identification of the health units as well as the identification of the medical procedures that can be performed by these health units. Cross-checking of these data sets requires substantial effort, indicating the need for an infrastructure to support the analysis proposed in this paper.

The databases referenced in Part A of Fig. 3 were the input for the modeling of a multidimensional database that allows the execution of queries in several dimensions. For this purpose, we used the resources of the SQL Server to perform data cross-checking from different sources (databases shown in Part A of Fig. 3). The result of this step was exported to a JSON file as shown in Part B of the same Figure. Fig. 4 presents a cube modeling scenario designed for analysis of approved medical billing invoices. The Table TB_AIH_PAGA_TABWIN_FATO is the cube's fact table to process the following indicators: Number of Approved Accounts (QTDE_APROV) and Value of the Approved Accounts (VALOR_APROV). The other tables are the dimension tables used to perform data cross-checking from the different sources. The table TB_PROCEDIMENTO was created from SIGTAP (referenced in Table 1). The table TB_AIH_PAGA_TABWIN was created from TABWIN. Finally, the tables TB_ESTABELECIMENTO and TB_MUNICIPIO were created from CNES (also referenced in Table 1).

The main architecture of the proposed visualization infrastructure is shown in Figure 5. It is comprised of a Tomcat Apache Web Server, a NoSQL server called CouchDB, the jQuery Ajax and the D3.js as a JavaScript library for manipulating DOM objects based upon data. The architecture is designed to implement parts B, C and D of the visualization pipeline represented in Fig. 3. The reason to select the NoSQL CouchDB database,

⁸ <https://goo.gl/LkkBPH>.

an acronym for Cluster of Unreliable Commodity Hardware ⁹, was that it is a distributed schemaless database that provides a ResTful JSON (JavaScript Object Notation) API to manage the databases and documents over HTTP that stores documents as JSON objects. This allows many data structures to be converted to and from JSON. Moreover, CouchDB's base HTTP engine easily scales up to handle heavy loads of requests, responding to HTTP requests using the GET, POST, PUT, or DELETE http methods to a specific CouchDB URL [12].

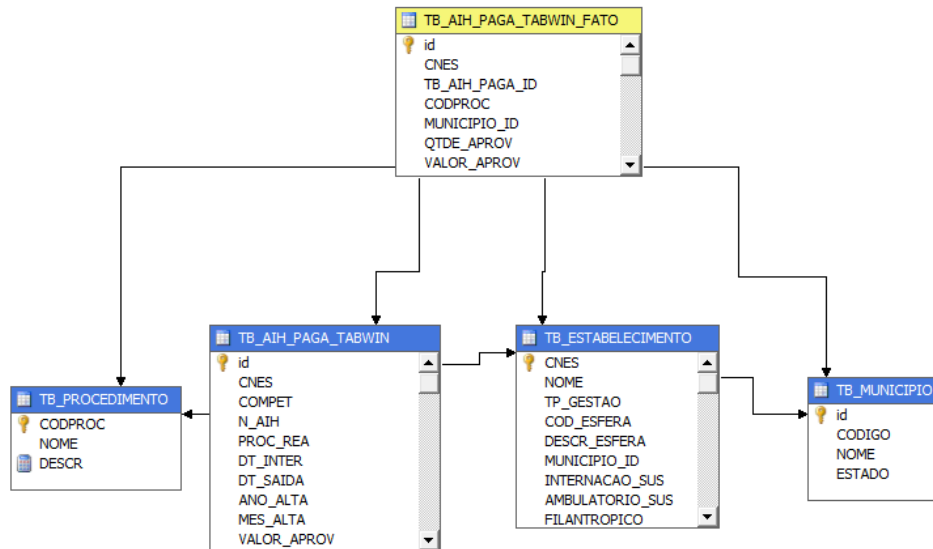


Fig. 4. A Cube Modeling Scenario for the Approved Medical Billing Invoices.

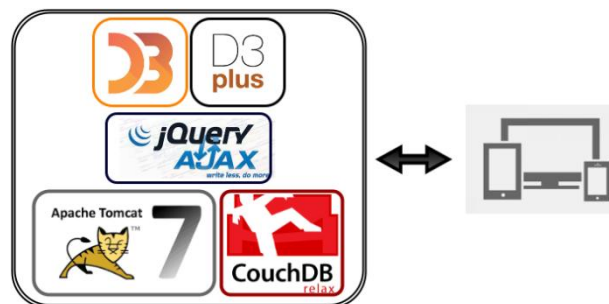


Fig. 5. The architecture of the proposed visual infrastructure.

```

1  [
2    {
3      "ESTADO": "SE",
4      "CNES": "2816210",
5      "COMPET": "201401",
6      "PROC_REA": "0413040240",
7      "Valor_APROV": 324.20,
8      "Qtde": 1}, ...
9  ]

```

Fig. 6. The data structure of a JSON file referenced in Part B of figure 3.

The text in Fig. 6 below is an excerpt of a JSON file that illustrates the representation of structured data from a health unit. This is an example of step B of the Visualization Reference Model from Fig. 2. The ID of this health unit is CNES 2816210 and it is located in Sergipe (SE) state. Of the remaining attributes: COMPET represents the year with the respective month, PROC_REA is the procedure type ID, Valor_APROV

⁹ <https://goo.gl/mtnk>.

is the amount of the approved account in Brazilian reais, and Qtde is the number of procedures performed by the health unit in the specific month. In the sequence, this data structure is the input to the assembly of the visual data structures represented in part C of Fig. 3. This new data structure maps data properties to visual properties as a requirement to render the visual metaphors. The result is that each visual metaphor can now represent the health data according to its characteristics.

The code segment shown in Fig. 7 is an excerpt of the real to visual attributes mapping for the treemap view. This mapping defines how data properties (in this case, health data attributes) are connected to visual attributes such as shapes, colors and positions on the visual metaphor. The `loadData()` function, starting on line 21, specifies the use of Ajax jQuery through the GET HTML command. The goal is to obtain data related to the values of approved accounts for each health unit from the Brazilian Northeast region. As can be seen on line 28, the CouchDB URL indicates the location of the JSON file containing data to be displayed in each view. In the example, the selected view is the treemap. The `loadData()` function maps the values from the JSON file to visual attributes that control the treemap visual metaphor using functions such as `.id()`, `.size()`, `.color()`. Once these values are passed to the view, the code calls the function `.draw()` to render the visual attributes in the visual metaphor. This reflects step D of the Visualization Reference Model. It is important to highlight that this step does not deal with rendering, but rather with building suitable data structures from which the views can be processed and rendered.

```

18  jQuery(function() {
19      loadData();
20  });
21  function loadData() {
22      var json;
23      jQuery.ajax({
24          type: 'GET',
25          xhrFields: {
26              withCredentials: false
27          },
28          url: "http://localhost:5984/tests/_design/vis/_view/treemap",
29          dataType: 'jsonp',
30          contentType: "application/javascript",
31          crossDomain: true,
32          async: false,
33          username: "xxx",
34          password: "xxx",
35          success: function(data){
36              json = data.rows[0].value;
37              console.log(json);
38              var root = nodes = json;
39              var visualization = d3plus.viz()
40                  .container("#viz")
41                  .data(json)
42                  .type("tree_map")
43                  .title("Contas Pagas")
44                  .title({"total": true})
45                  .id(["ESTADO", "COMPET", "CNES", "PROC_REA"])
46                  .size("QTDE_APROV")
47                  .color("Valor_APROV")
48                  .draw();
49      }

```

Fig. 7. Mapping real to visual attributes in the treemap view.

The last pipeline step is the view transformation used to instantiate the views (part D in Fig. 3) through the rendering of health data on the computer screen. At this moment, a specific visual scene is rendered on the computer screen.

The visual metaphors available in the infrastructure use the D3.js library. D3.js is an open source JavaScript library that provides resources for manipulating HTML documents based upon data, using JavaScript as the language for implementing the mapping of data to the documents [14]. D3 views are implemented in pure HTML and CSS, improving browser compatibility [15]. One of the benefits of D3 is how it exposes the designer directly to the web page and the corresponding browser existing technology

leveraged without having to use or create a whole new plotting language [16]. D3 allows binding data to a Document Object Model (DOM) that are originally converted at the browser end from HTML files sent through HTTP. DOM is in fact a tree structure, represented by hierarchical nodes where the top node being the main web-page or Document that can be adapted by Javascript. D3 aims at enabling manipulation of documents based on data, exposing capabilities of web standards such as HTML, SVG and CSS [16].

The first visual metaphor chosen to be used in the proposed environment is the treemap¹⁰. It represents large volumes of data as a hierarchy of recursively nested rectangles representing domain relevant entities also known as real attributes. Examples of these attributes are the number of approved accounts, approved amount, health unit ID's (CNES), among other attributes that can be visually represented by attributes like rectangle color and size.

The D3Plus extension was used to load the array of objects obtained in the query to CouchDB, by an AJAX call to the function that will shape the graph. In this function we define the keys of each treemap level and the variable to be associated with each visual attribute.

The second visual metaphor chosen to be used in the proposed environment is the sunburst¹¹. It is a radial space-filling visualization technique for displaying tree like structures that are laid out radially [16]. The root node is shown at the center of the display. Successive levels of nodes with their respective attributes and are drawn further away from the center. Each level is given equal width, but the circular angle swept out by a node directly corresponds to its size [17]. Thus, the size of any two nodes in the hierarchy can be compared via the angles they subtend. The absolute area of two "pie slices" can be used only to compare sizes on the same level [17]. An example of this visualization is shown in Figure 8.

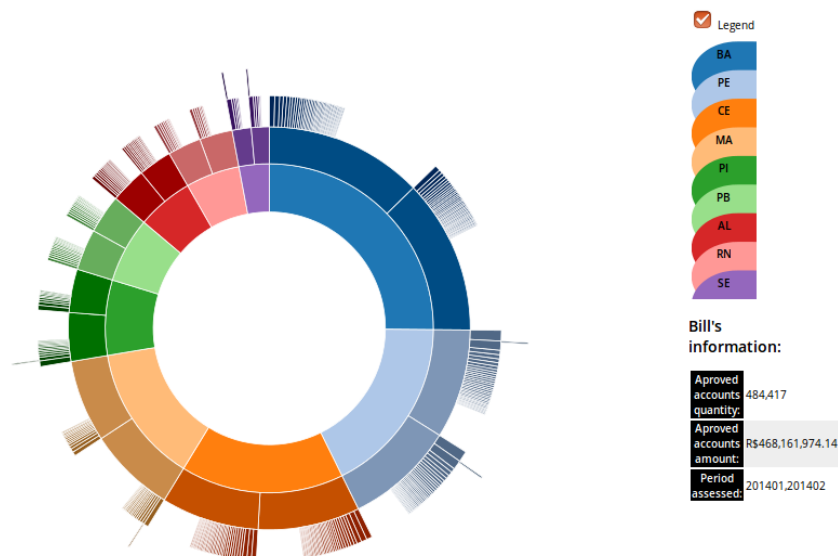


Fig. 8. A Panoramic view of approved accounts in sunburst (first level on states).

5. The Rationale behind the Visual Metaphors to Support the Analysis

The goal of the data processing and arrangement as performed by the cube modeling represented in Fig. 3 and 4 is to support the data analysis in levels as set out below. A cube, as an interactive analysis session, represents a collection of events of a particular business process over the lifetime of a set of databases. The collection may contain a large number of events—thousands, Million, or even billions—but the events are not presented individually [18]. Fig. 9 conveys the four phases of the strategy analysis associated with different levels of data arrangement. As far as the user makes progress in the analysis, he/she interacts with

¹⁰ <https://en.wikipedia.org/wiki/Treemapping>.

¹¹ <http://www.cc.gatech.edu/gvu/ii/sunburst>.

the visual metaphors treemap and sunburst. In the case of the treemap, the user performs the semantic zooming. the transition from Level 1 to Level 4 of the visual data structure characterizes the semantic zooming. This transition can occur both backwards and forwards.

This type of zoom transforms the view to the underlying meaning inside the target object. Usually this mechanism reaches into the derived data contained in the data table stage [19]. In the case of the sunburst, the user selects the arcs to highlight the data to be presented in the canvas.

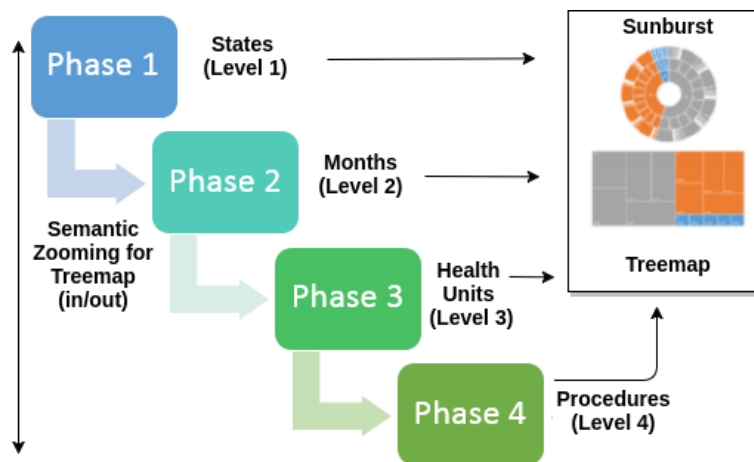


Fig. 9. Proposed strategy to analyze health data.

To feed both visual metaphors in the first level, the visual data structure is comprised of several arrays of objects as presented in Fig. 10.

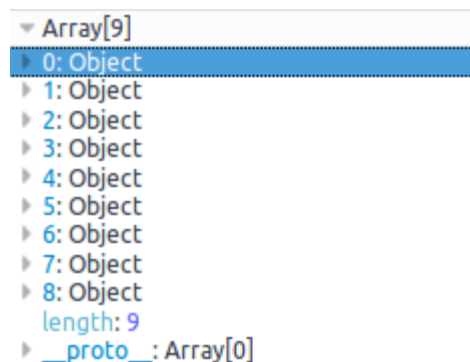


Fig. 10. The visual data structure to feed the visual metaphors (Part C of figure 3)

The first phase presented in Fig. 8 and 11 shows data focused on the Level 1 and organized by Brazilian states (where *Ba* stands for Bahia, *Pe* for Pernambuco, *Ma* for Maranhao, *Ce* for Ceara, *Pi* for Piaui, *Pb* for Paraiba, *Al* for Alagoas, *Rn* for Rio Grande do Norte and *Se* for Sergipe) in the sunburst and treemap views. At this point, the user can have a broad view of both the number of approved accounts and the respective amount paid for them in the case of the treemap view. However, in this phase, the treemap view shows the distribution of number of approved accounts for each state (area of each rectangle) and total value approved for each state (using the colors shown in the heat map legend below the figure). In contrast, the sunburst view shows the number of approved accounts organized in a hierarchy of levels (States, Months, Health Units, and Procedures) simultaneously.

In the treemap view, the transition from the first to the second level (from Fig. 11 to Fig. 13) occurs when the user clicks in one of the rectangles representing one of the states. This semantic zooming causes the interface to show data from the chosen state (here, Pernambuco) organized by months. In the case of the sunburst view, the transition from the first to the second level (from Fig. 14 to Fig. 15) occurs when the user clicks in one of the arcs representing one of the states. Fig. 12 shows the visual data structure that drives both visualizations. For example, the *key* value is "PE", representing Pernambuco, where the value "85090" in *Qtde* corresponds to the number of approved accounts for Pernambuco in the constant time interval for the analysis.

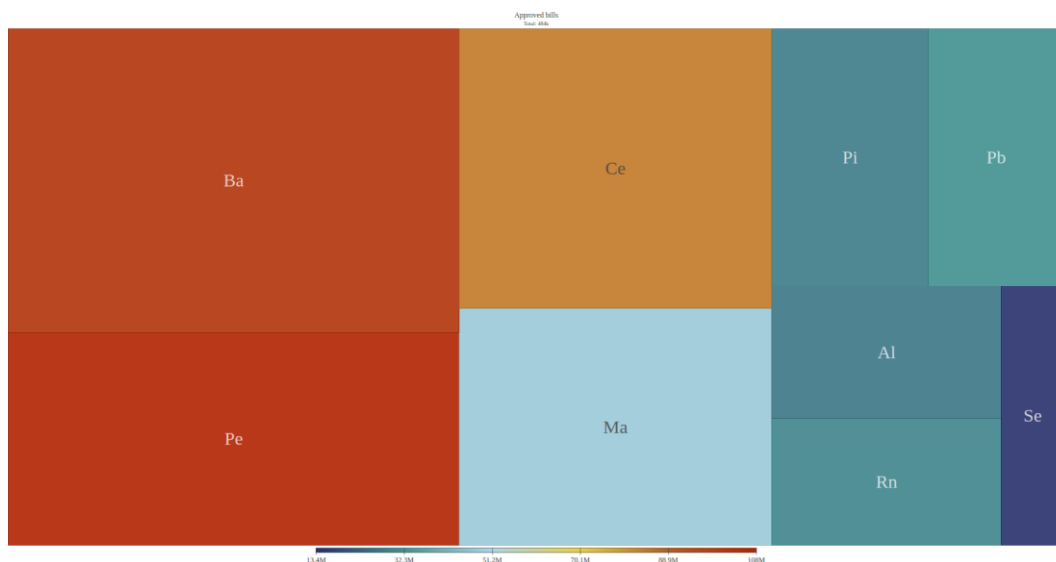


Fig. 11. Approved accounts and budget allocation per states in treemap.

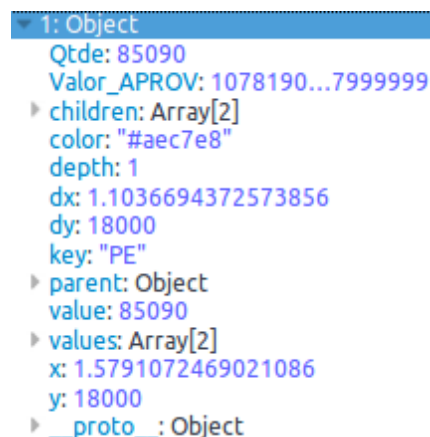


Fig. 12. The visual data structure for the first level (phase 1 of figure 9).

The second level presented in Fig. 13 and 15 for the treemap and sunburst, respectively, shows data organized by months. In the case of the treemap view, the transition from the second to the third level (from Fig. 13 to Fig. 16) occurs when the user clicks in one of the rectangles representing one of the Months. This action zooms the interface to visualize the data organized by health units. The same basic change happens in the sunburst view, (from Fig. 15 to Fig. 19) when the user clicks in the arc for the month of February 2014. The visual data structure presented in Fig. 17 is used by both visual metaphors. For example, the key value is "201402", representing February of 2014, whereas the value "42523" in value corresponds to the number of approved accounts in the selected month.

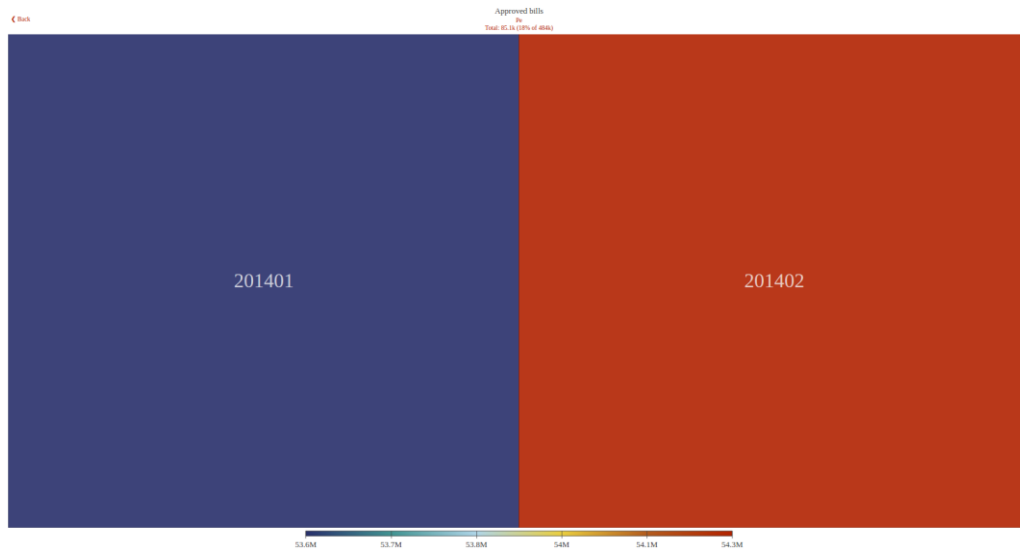


Fig. 13. Approved accounts in treemap (second level on months).

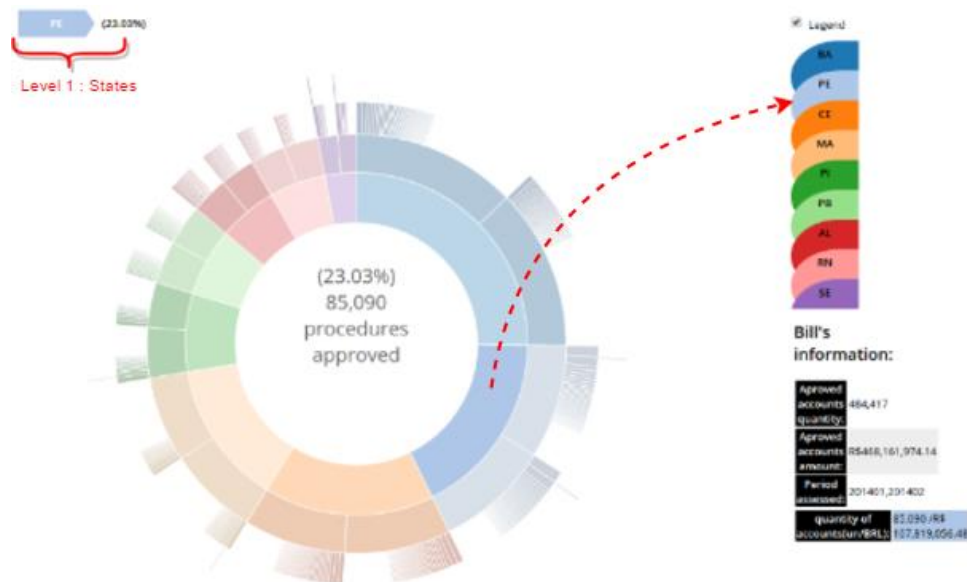


Fig. 14. Selecting pernambuco state.

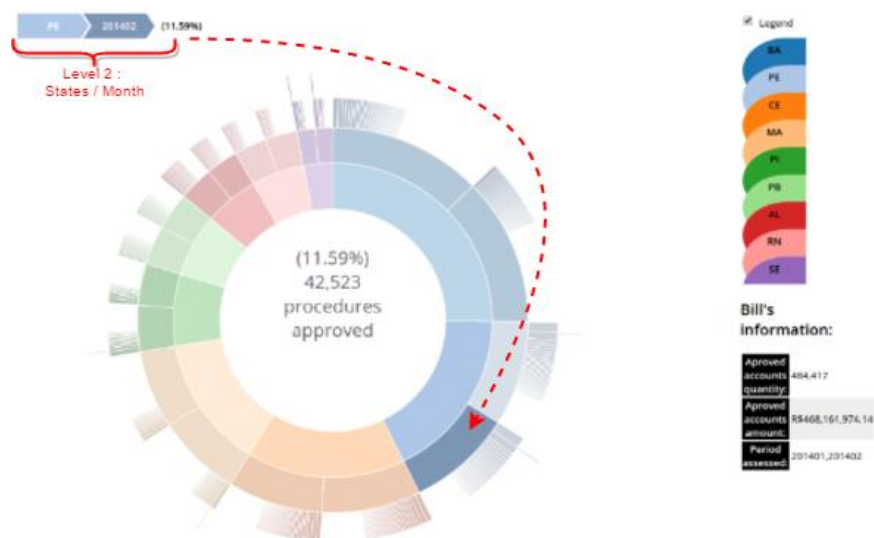


Fig. 15. Selecting february/2014 for Pernambuco state.

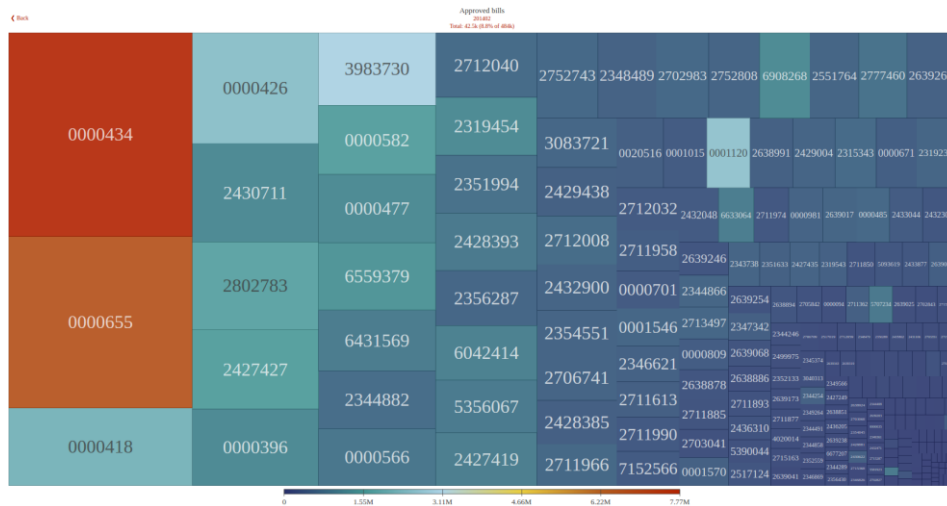


Fig. 16. Approved accounts in treemap (third level on health units).

```

1: Object
  Qtde: 85090
  Valor_APROV: 1078190...7999999
  children: Array[2]
    0: Object
      1: Object
        Qtde: 42523
        Valor_APROV: 542555...999999
        children: Array[235]
          color: "#a9c8da"
          depth: 2
          dx: 0.5515493651486169
          dy: 18000
          key: "201402"
          parent: Object
            value: 42523
          values: Array[235]
            x: 2.1312273190108773
            y: 36000
          __proto__: Object
    1: Object
      Qtde: 42523
      Valor_APROV: 542555...999999
      children: Array[235]
        color: "#a9c8da"
        depth: 2
        dx: 0.5515493651486169
        dy: 18000
        key: "201402"
        parent: Object
          value: 42523
        values: Array[235]
          x: 2.1312273190108773
          y: 36000
        __proto__: Object
  __proto__: Object

```

Fig. 17. The visual data structure for the second level (phase 2 of Fig. 9).

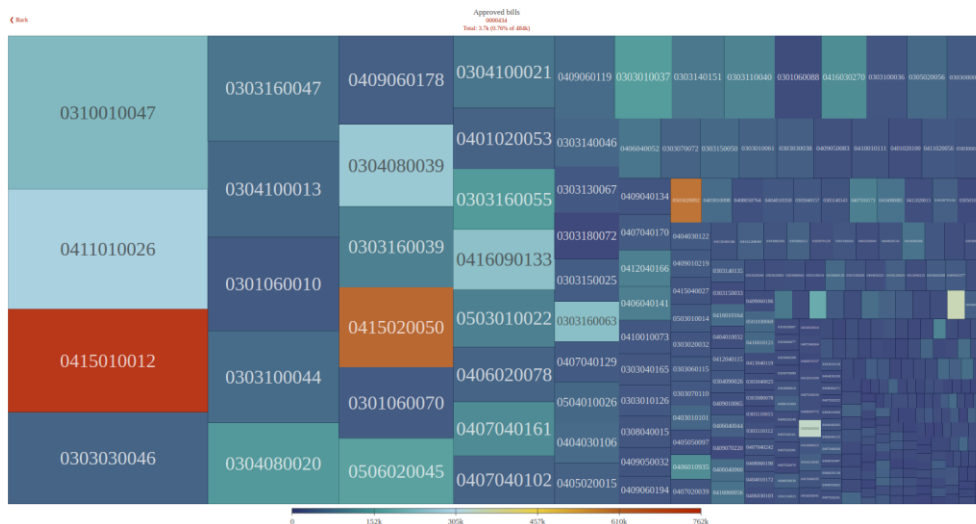


Fig. 18. Approved accounts in treemap (fourth level on procedures).

The third level portrays data of health units for a specific state and month. In treemap view, the user can click on a health unit's rectangle in order to perform semantic zooming and see the data organized by procedure.. In the sunburst view, this is only possible for specific cases, due to restrictions in the granularity of the data. The result is that not all health units of the selected state have their respective procedures visually represented because in most cases they correspond to less than 0.1% of the overall accounts. Given

this percentage, in the sunburst view it is not possible to show the corresponding arcs of the fourth level (the procedures) due to the limitation in the visual metaphor to visually present small arcs. However, in the case of Fig. 20, the procedure "Cerebrovascular accident treatment/Ischemic or Sharp bleeding" (identified as 0303040149), which corresponds to 0.18%, is able to be presented. The visual data structure represented in Fig. 21 shows how data is organized up to level 3.

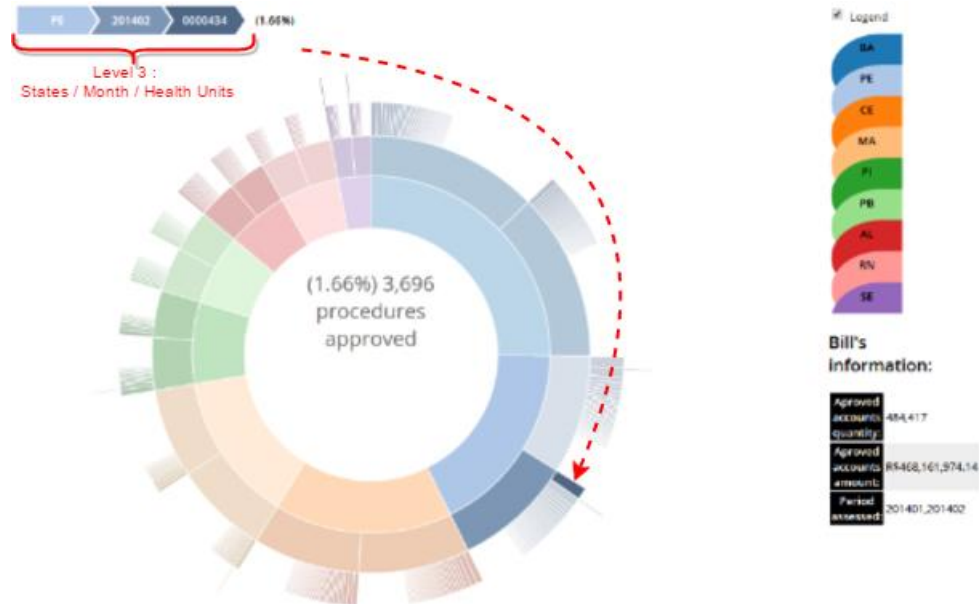


Fig. 19. Selecting a health unit in pernambuco state with data from February/2014.

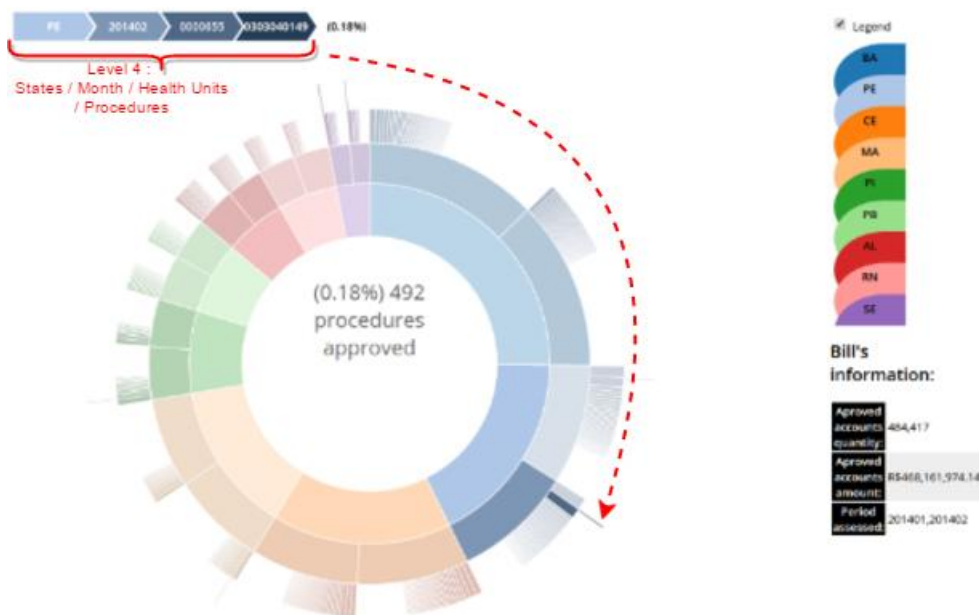


Fig. 20. Selecting a procedure performed in pernambuco state with data from February/2014 and Health Unit 0000434.

The fourth level visually represents data of procedures performed by a health unit from a state in a specific month. Table 3 presents a comparison between features and resources provided by treemap and sunburst to support health data comprehension.

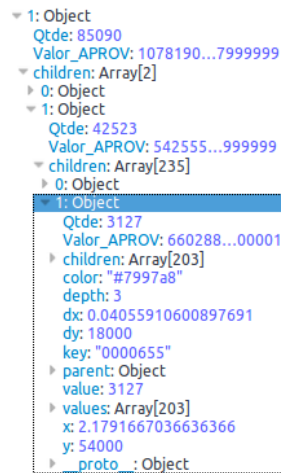


Fig. 21. The visual data structure for the third level (phase 3 of Fig. 9).

Table 3. Treemap Versus Sunburst Data Comprehension Support

Features and Resources	Treemap	Sunburst
Support the visual analysis of a set of attributes	Yes (Amount of Accounts and Budget Allocation)	Limited to the Amount of Accounts. The value is available in the frame.
Use of Visual Attributes	Rectangle Area representing the Amount of Accounts; Color representing the Value of the Accounts	Arcs (size of the slices) representing the amount of accounts; Color representing the share among States
Levels	1st level - States 2nd level - Months 3rd level - Health Units 4th level -Procedures Each level can be reached moving from the current to the next or previous one using the Semantic zooming. For example, after selecting a State in the first level the user can reach the months in the second level.	Each concentric arc represents a level. From the inner concentric arc to the outer the following entities are represented: State, Month, Health Unit and Procedure.
Multilevel versus Specific Level Data Visualization	Allows analysis of data from a specific level.	Allows the analysis of data from different levels.
Data Level Granularity	Data granularity is aligned with each level. The view provides a clear representation of data until the fourth level.	Data granularity is aligned with the set of the four levels. The analysis of the third and fourth level is hindered, thus the user has difficulty to analyze data related to health units and procedures.

6. Using the Visual Approach to Analyze the Data

As we have mentioned in Section 2, the Brazilian Northeast region has suffered from shortage of public medical care provided by the Brazilian Public Unified Health System (SUS). For this reason, we decided to focus this example of use on data related to approved accounts of health units from this region. The data used as starting point in this example is publicly available by the Brazilian Ministry of Health ¹².

6.1. Example of Use

This example of use intends to answer the following questions: **Question 1 (Q1):** *What are the top three*

¹² <https://goo.gl/LkkBPH>.

Brazilian Northeast states in terms of highest approved values for performed procedures? The identification of these states is important to characterize how Brazilian states have provided support for citizens in terms of Health assistance and correlate it in terms with their respective population size. **Question 2 (Q2):** *Considering the states selected from Q1, which month had the highest approved value for performed procedures?* Selecting the month with the highest value for performed procedure is a possible strategy for a target sample to analyze health units data. **Question 3 (Q3):** *Considering the month indicated by Q2, what are the top three corresponding health units of each state selected from Q1 that received the highest approved values for performed procedures?* The identification of these health units has the goal to characterize their profile and to identify what can be shared with other health units in terms of lessons learnt. **Question 4 (Q4):** *What are the top three procedures that most contributed to these highest values?* It is important to know the list of procedures that demand more budget allocation in health units to indicate possible focus on health precautions targeting these diseases.

6.2. The Target Data

To answers these questions, we combined the use of the treemap and sunburst views whose rationale was described in the previous section. Both views effectively represent hierarchical data and can be used complementing one another.

An important decision of the visual analysis is how to use the available visual metaphors. Considering the strategy proposed in Fig. 9, the panoramic view provided by treemap (Fig. 11) can initially represent data related to the first level, i.e. the states in terms of approved accounts and the respective budget allocated for them. On the other hand, the sunburst visual metaphor not only provides a panoramic view of data (Fig. 8), but it also represents the subsequent levels at the same time. This helps the user to know to which extent the set of analyzed states, period of time (in this case, month), and health unit contribute in the amount of approved accounts.

6.3. Results and Data Analysis

6.3.1. Question 1 (Q1)

To answer **Question 1 (Q1)**, we identified the Brazilian Northeast states that had the highest approved values for performed procedures. In this example of use, we analyzed data collected from January to February 2014. Similar analysis can be performed with more months in the data set. In this version of treemap, the rectangle visual attribute is associated with the amount of approved accounts of a given state. The color visual attribute of each rectangle is associated with the budget allocated for the respective accounts. These values are represented along with the respective colors as a scale (Fig. 22). As can be seen in the treemap view shown in Fig. 11, the top three states in terms of the highest values for performed procedures are Pernambuco (85090 approved accounts and the corresponding value of 107.8 Million of Reais paid for them), followed by Bahia (122.000 approved accounts with 102 Million of Reais) and finally Ceara (77.800 approved accounts with 83.9 Million of Reais). The sunburst view in Figure 8 portrays the states in arcs whose size is directly proportional to the number of approved accounts. Hovering the mouse over the top three states in terms of approved accounts, we can conclude that they are also the top three in terms of approved values for performed procedures except for the order. Bahia has the second highest value while Pernambuco has the highest value. These values can be checked in the legend of the sunburst view when the hovering the mouse to select the state.



Fig. 22. Color scale in treemap.

Considering all the states from the Northeast region, the approved values for performed procedures range from R\$ 13.4 to R\$ 108 Million of Brazilian Reais (Figure 22). Comparing these values with the population size of each state ¹³, we can conclude that Pernambuco, despite having fewer performed procedures than Bahia, received more funding. Considering that Pernambuco performed a given number of procedures in the analyzed month, the final results was a higher budget allocation than those performed by Bahia. The Table 4 illustrates this scenario.

Analyzing again Fig. 11, we observed that Ceara and Maranhao states are represented in similar rectangle areas and their respective colors are quite different. These two states, despite having close number of approved accounts, received different values for the accounts. Ceara received for 77.798 approved accounts the value of R\$ 83.930 Million of Brazilian Reais and Maranhao received for 65879 approved accounts the value of R\$ 47.452 Million of Brazilian Reais. States with the lowest number of approved accounts are represented in the lower right vertex of the treemap visual metaphor. These states are Alagoas (AL), Rio Grande do Norte (RN) and Sergipe (SE). When the user select a specific state (treemap rectangle), a text box portrays data related with the selected state to support the analysis.

Table 4. Top Three Brazilian Northeast States with Highest Approved Values

State	Approved Value (Million of Reais - R\$)	Approved Accounts	Population Size (Million Inhabitants - 2014)
PE	107.82	85090	9.2
BA	102.09	121745	15.1
CE	83.93	77798	8.8

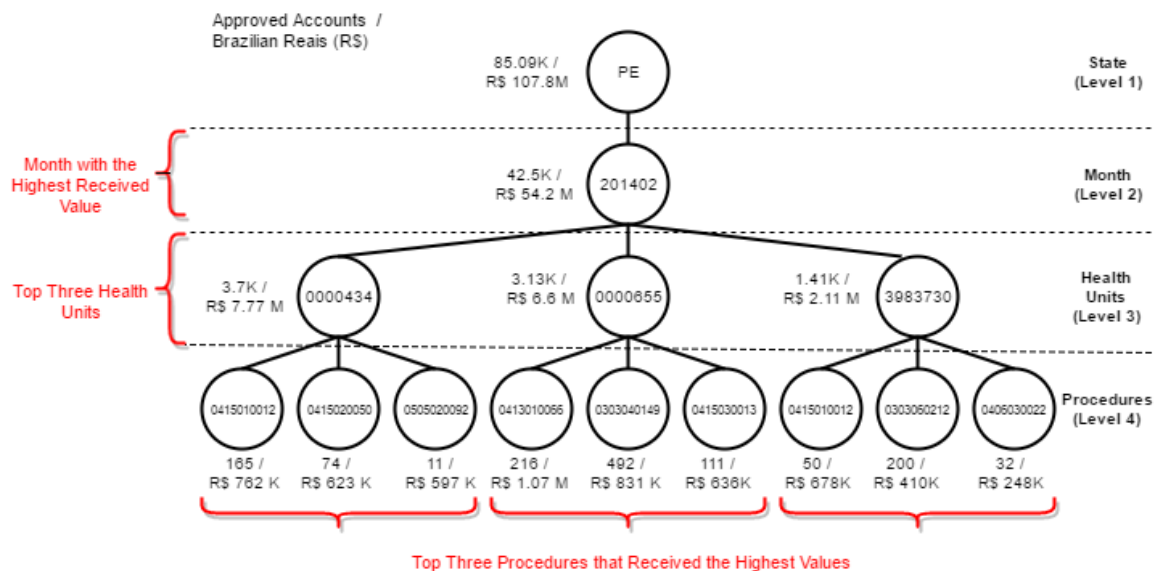


Fig. 23. Answers for questions Q2, Q3 and Q4 for the pernambuco state.

6.3.2. Question 2 (Q2)

To answer **Question 2 (Q2)**, we considered the top three states from Question Q1 and moved to phase 2 (level 2) of the proposed strategy to analyze health data, as described in Figure 9, using both the treemap and sunburst view. In the case of the treemap view, we clicked in each rectangle that represents the states

¹³ <http://www.ibge.gov.br/apps/populacao/projecao/>.

selected from Question 1 (Q1) to identify the month that had the highest value for performed procedures. Fig. 23, 24 and 25 portray in the level 2 the respective months to answer Q2 for each state: month 20142 for Pernambuco and month 201401 for Bahia and Ceara. In the case of the sunburst view, we just hover the mouse over the arcs of the second level of the selected states to identify the months.

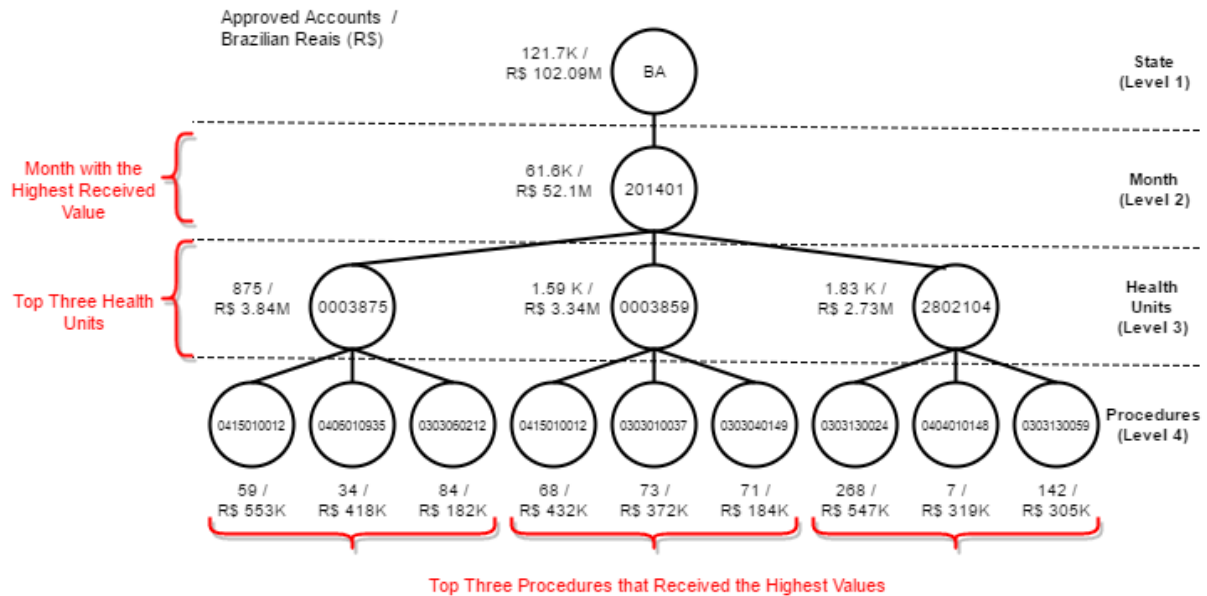


Fig. 24. Answers for questions Q2, Q3 and Q4 for the Bahia state.

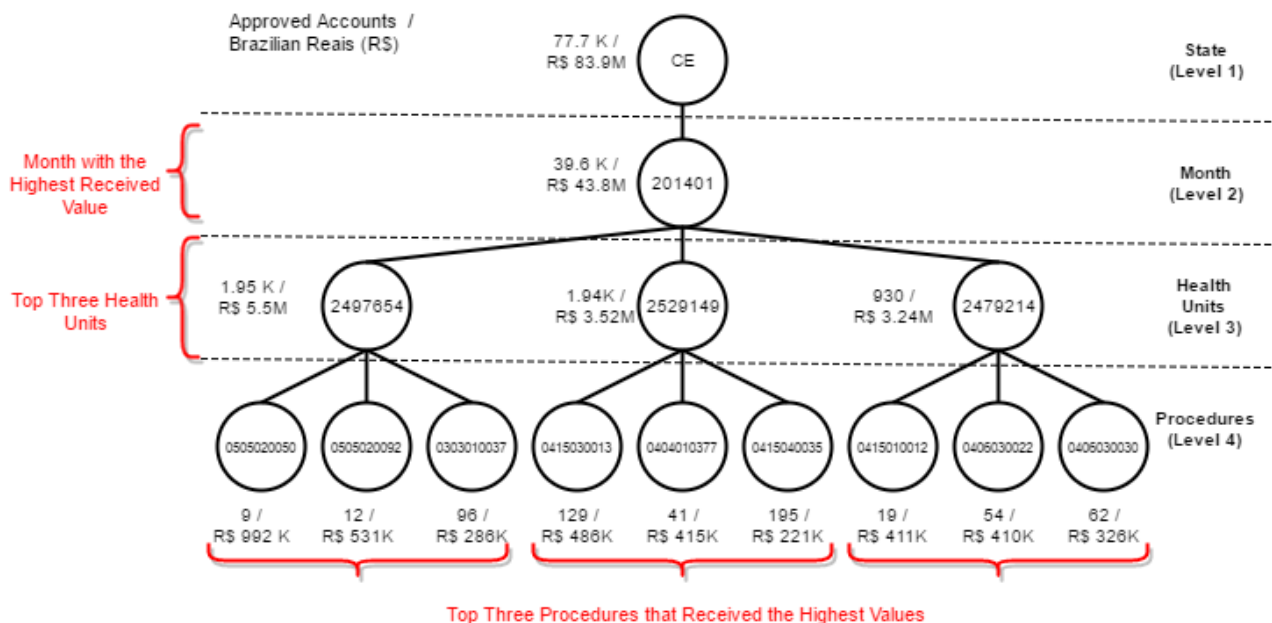


Fig. 25. Answers for questions Q2, Q3 and Q4 for the Ceara state.

6.3.3. Question 3 (Q3)

To answer **Question 3 (Q3)**, we considered the months to answer **Question 2 (Q2)** and the states to answer **Question 1 (Q1)** obtained for Q1 to identify the top three health units of each state that received the highest approved values for performed procedures. To accomplish this, we moved to phase 3 (level 3)

of the proposed strategy to analyze health data, as described in Fig. 9. We followed the same rationale regarding the visual attributes of treemap, where the rectangle area is associated with the amount of accounts and the color is related to the value of the accounts. Considering the visual scenario presented in Fig. 16, the top three health units that received the highest values for performed procedures for Pernambuco, Bahia and Ceara are presented the level 3 in Fig. 23, 24, 25. Using the sunburst view (Figure 19), we can hover the mouse over the arc of the third level after each month of the selected states to identify the months.

6.3.4. Question 4 (Q4)

To answer **Question 4 (Q4)**, we searched for procedures that most contributed to the highest values of approved accounts for the selected months as indicated before. Considering the visual scenario presented in Figure 18, the top three procedures that most contributed to the highest values of approved accounts for the selected months for Pernambuco, Bahia and Ceara are presented the level 4 in Fig. 23, 24, 25. Using the sunburst view (Fig. 19), we can hover the mouse over the arc of the fourth level after each month of the selected states to identify the months. However, due to limitations of the sunburst view, we cannot spot these procedures using this view.

From the data obtained to answer **Question 4 (Q4)** and considering what was presented in Table 5, we concluded that the three Health Units that received the highest values were 0000434 and 0000655 from the Pernambuco state with R\$ 7.77 and R\$ 6.6 Million Reais of budget allocation respectively. Moreover, the Health Unit 0000434 in Ceara state highlighted by the R\$ 7.77 Million Reais of budget allocation.

Table 5. Top Procedures Distribution Per State

Procedures (Identification/Description)	States		
	PE	BA	CE
0303010037 - Treatment of other bacterial disease		X	X
0303040149 - Cerebrovascular accident treatment (Ischemic or Sharp bleeding)	X	X	
0303060212 - Heart failure treatment	X	X	
0303130024 - Cardiovascular Diseases long-term care patient treatment		X	
0303130059 - Neurological Diseases long-term care patient treatment		X	
0404010148 - Cochlear implant		X	
0404010377 - Tracheostomy			X
0406010935 - Myocardium re-vascularization with extracorporeal circulation (with two or more grafts)		X	
0406030022 - Coronary Angioplasty with the implantation of stents	X		
0406030022 - Coronary angioplasty with implantation of two stents			X
0406030030 - Coronary angioplasty with implantation of stents			X
0413010066 - Treatment of burns	X		
0415010012 - Treatment with multiple surgeries	X	X	X
0415020050 - Oncology sequential procedures	X		
0415030013 - Surgical treatment in polytraumatized	X		X
0415040035 - Ulcerate debridement of devitalized tissues			X
0505020050 -Liver transplant (Organ from deceased donor)			X
0505020092 - Kidney transplant (Organ from deceased donor)	X		X

All the health units represented in Fig. 23, 24 and 25 are accredited by the Brazilian Ministry of Health to perform high complexity procedures such as treatment with multiple surgeries (0415010012). This treatment has a higher value than other procedures. The health units identified by 3983730 and 2479214 are highlighted for services provided by heart treatments with high costs. Another interesting item of information is that from the set of the selected health units related to the answers of **Questions (Q1-Q4)** from Ceara are coordinated by the city government while the others from Pernambuco and Bahia are

coordinated by the state governments.

7. Conclusion

This paper presented an infrastructure to support the analysis of data from the Brazilian Public Unified Health system based on interactive visual resources provided by two visual metaphors: treemap and sunburst. The analysis focused on data related to approved accounts of health units from the Brazilian Northeast region due to reports of shortage of public medical care in this region. To perform the analysis, we proposed a strategy that goes through four phases/levels. Each of these phases/levels focus on key entities of the analyzed data: States, Months, Health Units and procedures. The treemap and sunburst views were implemented to enable the analysis of these key entities in an interactive way using resources such as the semantic zooming. The analysis had the goal to answer four questions considering the proposed strategy. As a result of this work, we could identify states that stood out due to the budget allocation for services provided in a specific month and the respective health units and procedures that contributed to this result. Throughout the work we could also identify research gaps that need further investigation as follows: i) improvement in the treemap view to portray a panoramic view rather than a level specific data representation; ii) improvement in the sunburst view to portray data with low granularity through the use of the zoomable sunburst; iii) inclusion of filters in the view to enable the user to select a specific data subset to be shown on the canvas.

Acknowledgment

We thank FAPESB (State of Bahia Research Foundation) for providing the scholarship (grant #BOL1258/2016) to the third author of this paper.

References

- [1] Mays, G. P., Scutchfield, F. D., & Bhandari, S. A. S. M. W. Understanding the organization of public health delivery systems: An empirical typology. *Milbank Quarterly*, 88, 81-111.
- [2] Scutchfield, F. D., Bhandari, M. W., Lawhorn, N. A., Lamberth, C. D., & Ingram, R. C. (2009). Public health performance. *American Journal of Preventive Medicine*, 36, 266-272.
- [3] Barros, B A. J. D., & Bertoldi, A. D. (2008). Out-of-pocket health expenditure in a population covered by the family health program in brazil. *International Journal of Epidemiology*, 37, 758-765.
- [4] Lavrač, N. *et al.*. (2007). Data mining and visualization for decision support and modeling of public health-care resources. *Journal of Biomedical Informatics*.
- [5] Santos, R. S., *et al.* (2013). A data mining system for providing analytical information on brain tumors to public health decision makers. *Computer Methods and Programs in Biomedicine*.
- [6] Baldonado, W. A. M. Q. A., & Woodruff, A. K. (2000). Guidelines for using multiple views in information visualization. *Proceedings of the Working Conference on Advanced Visual Interfaces*.
- [7] Spence, R. (2001). *Information Visualization*.
- [8] Godfrey, P., Gryz, J., & Lasek, P. (2016). Interactive visualization of large data sets. *IEEE Transactions on Knowledge and Data Engineering*, 28, 2142-2157.
- [9] Munzner, T. (2014). *Visualization Analysis and Design*.
- [10] Yi, J. S., Kang, Y. A., & Stasko, J. (2007). Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics*.
- [11] Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). Readings in information visualization: Using vision to think, *Morgan Kaufmann*.
- [12] Silva, A. N., Carneiro, G. D. F., & David, J. M. N. (2014). Towards the development of a framework for

multiple view interactive environments. *Proceedings of the 2014 11th International Conference on Information Technology*.

- [13] Anderson, J. C., Lehnardt, J., & Slater, N. (2010). *CouchDB: The Definitive Guide*.
- [14] Heydt, M. (2015). *By Example*, Packt Publishing Ltd
- [15] Bostock, M., Ogievetsky, V., & Heer, J. (2011). D3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*.
- [16] Amre, T., & Stamboliyska, R. (2016). *Getting Started with D3*.
- [17] Stasko, J., Catrambone, R., & Guzdial, M. (2000). McDonald, an evaluation of space-filling information visualizations for depicting hierarchical structures. *International Journal of Human-Computer Studies*.
- [18] Back, D. W., Goodman, N., & Hyde, J. (2013). *Mondrian in action: Open source business analytics*. Manning Publications Co.
- [19] Boulos, M. N. K. (2003). The use of interactive graphical maps for browsing medical/health Internet information resources. *International Journal of Health Geographics*.



Carlos Helano Aquino do Nascimento was born in Fortaleza, Ceará, Brazil in 1976. He has a MSc in computational modeling from SENAI/CIMATEC-BA (2012) and a BSc in mathematics from Universidade Salvador – (UNIFACS/Brazil - 2001). He is currently a computer science lecturer and researcher at UNIFACS in the area of computing and has experience in development of software. His research interests are related to data analysis, data visualization and data mining.



Glauco de Figueiredo Carneiro holds a PhD in computer science from the Federal University of Bahia (UFBA/Brazil). He is currently a professor of computer science at Universidade Salvador (UNIFACS/Brazil). His research interests include topics related to information and software visualization, cloud computing, machine learning and empirical software engineering.



Ed Santana Martins Costa is an undergraduate student of the Computer Engineering Program from Universidade Salvador (UNIFACS/Brazil). He has currently a student fellowship from the Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB). His research interests are related to data visualization and software testing.



Ethan V. Munson was born in Boston, Massachusetts, USA in 1956. He received a B.A. in psychology (1978) and a B.A. in computer science (1986) from the University of California, San Diego. He received the M.S. (1989) and Ph.D. (1994) in computer science from the University of California, Berkeley. He is currently professor of computer science and associate dean for academic affairs in the College of Engineering and Applied Science at the University of Wisconsin-Milwaukee, where he is also the Director of the Center for Advanced Embedded Systems. He has also been a visiting professor at Université Pierre Mendès-France and a visiting researcher at the Universidade de São Paulo. His current research focuses on configuration management for software and for documents and on human-computer interaction. Previous research emphasized multimedia authoring, especially the design and use of style sheet languages. Dr. Munson is a member of the American Society for Engineering Education and the Sociedade de Computação Brasileira. He is a senior member of ACM and was Chair of ACM SIGWEB from 2006 to 2011. Dr. Munson is a recipient of a National Science Foundation CAREER award, as well as four NSF educational grants, and a variety of industrial funding. He has had extensive conference organization activities, especially with the ACM Symposium on Document Engineering and ACM Hypertext.