# Information visualization

Sergio Jose Ruiz Sainz*
1379550
Eindhoven University of Technology

Dagoberto Jose Herrera Murillo†
1379291
Eindhoven University of Technology

**ABSTRACT**

This report describes the "Argentinian National Budget 2015-2019" application, which is designed so that through it any citizen can visualize the structure of the national budget, paying special attention to the functions of spending (social services, defense, debt, etc.). The principles of visualization presented here apply to the visualization of public accounts in any country in the world that uses a multidimensional financial model.

**Index Terms:** Visualization—Open Budget—Interaction—Bubble charts;

## 1 INTRODUCTION

Budget data could be defined as detailed records about spending and revenue. This information is the clearest manifestation of a government's priorities. Understanding these priorities in detail is of the utmost importance to guarantee the efficiency and transparency of the public function. However, in most cases there are no analysis and visualization tools to help citizens and other relevant stakeholders scrutinize how resources are spent and detect possible risks of mismanagement.

Budget design is a process in which multiple actors with different and frequently competing interests participate. This diversity of viewpoints makes the data generated in the budgeting process inherently complex and multidimensional. Each dimension of the data serves the needs of different users. Jacobs, Helis, & Bouley (2009) [4] proposed the concept of Multidimensional Modeling of Fiscal Information (MMFI) [3], in which each component of a public budget is associated to at least four hierarchical dimensions: administrative, economic, functional, and chronological. The administrative dimension refers to the area of government that is responsible for using the resources (e.g. Supreme Court of Justice). The economic dimension contains the type of expenditure incurred (e.g. salaries). The functional dimension identifies the purpose of spending (e.g. education). Finally, the geographic dimension points to the regions where income comes or where spending is incurred. The components of a public budget are monetary amounts, which may correspond to the different phases of the budget execution process (e.g. approved, adjusted, executed, committed, paid).

Responding to the need of visualizing this specific type of complex data, we build an interactive application to visualize government budget data with the techniques studied in class.

## 2 PROBLEM DESCRIPTION AND TASK ANALYSIS

Data published by central government, local authorities and public bodies is a strategic asset which can encourage citizen participation. When presented clearly, it has an enormous potential for increasing transparency, reinforcing communication and improving accountability. In such manner, people can monitor what their governments have done or fail to do. Nevertheless, datasets are sometimes presented as large, multidimensional, highly redundant and not well-structured, especially in terms of terminology.

The aim of this project is to provide an analytical tool to empower citizens with data, so they can easily see and understand complex budget information through intuitive visualizations. As follows, more meaningful answers can be provided to questions such as how much money is spent on each budget item, who decides how much funding is allocated for each budget item.

To accomplish the previous objective, a public dataset generated by the government of Argentina was utilized: `https://www.presupuestoabierto.gob.ar`. This dataset summarizes the distribution of the budget, of the central government, and its cumulative annual execution to the date. For this project, a total of 10,278 budget items which correspond to the period 2015-2019 were analyzed.

The most relevant questions that could be asked for a person to understand the operation of a public budget are the following:

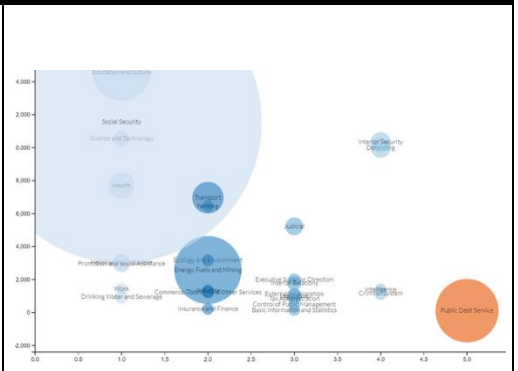
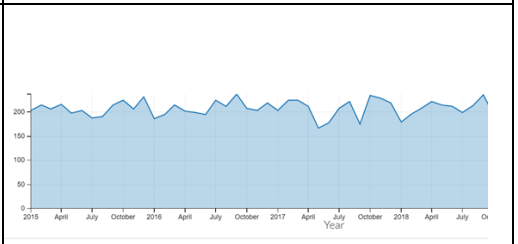
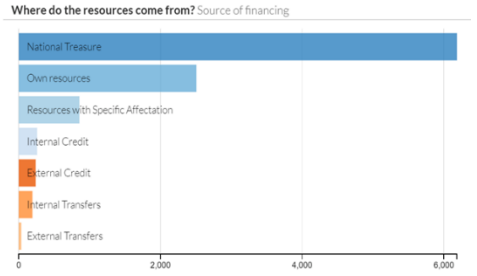
- How the public budget is spent in terms of purpose?
- How has the budget been spent over time?
- Where the resources come from?
- To which administrative branch does the expense correspond?
- Where is the money spent?
- Who spend the money?
- What does each budget component consist of?

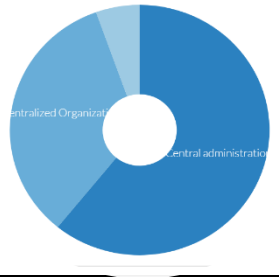
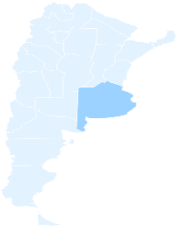
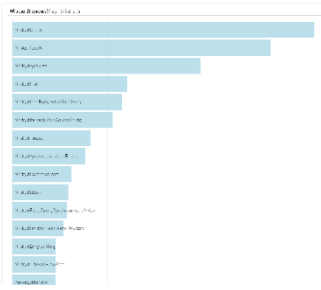
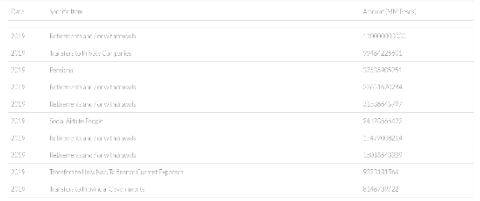
## 3 VISUALIZATION DESIGN

Our application structures the visualization of a budget information dataset linearly, following the concepts of the profile page and the inverted pyramid (writing style is characterized by putting the most relevant information at the beginning to grab the reader's attention from the beginning). As already mentioned, the multidimensional nature of the budget information causes different stakeholders to focus their attention on different questions. This is vital because there may be multiple possible design structures and the decision on which to use depends on the profile of the main user. Our analysis assumes that the main stakeholder is a citizen who has a special interest in understanding the purpose and function of the expense. That is why the head of our application focuses first on addressing this issue. From this central question, in the successive views you can explore elements of other dimensions in the form of basic 'w-questions' (When? Where? Who? What?). Table 1. contains a summary of the design decisions that have to do with the selection of the most appropriate visualization techniques to answer the questions of our use case, as well as a description of the visual encoding used to represent each one of them.

*e-mail: s.j.ruiz.sainz@student.tue.nl
†e-mail: d.j.herrera.murillo@student.tue.nl

Table 1. Design choices

| Question | Visualization technique | Visual encoding | Overview |
|---|---|---|---|
| How the public budget is spent in terms of purpose? | Understanding the purpose of spending is the main objective of the entire analysis. Each of the purposes is simultaneously associated with categorical variables (type of expense: debt vs. non-debt) and numerical variables such as the amount of the budget and the expenditure segment. Therefore, to represent these attributes in the same visualization we require a more sophisticated multidimensional technique than that offered by simple alternatives such as bar and pie charts. In this context the bubble chart turns out to be a useful tool to jointly represent several categorical and numeric dimensions simultaneously. The use of a sunburst or tree-maps was also valued for nesting categories of proportions. However, these procedures would leave out the numerical dimensions. The advantage of the bubble chart is that it can encode hundreds of data points into a limited area. However, it is less perceptually accurate than other representations like bar charts. | In this case, each **expenditure functionality** (29 categories) is coded as a circle, the radius of the circle represents the **total budget** (millions of Argentine pesos) that corresponds to said item. On the x-axis the **section of the expense** to which each functionality belongs (Value 1-5) is considered, while on the y-axis the **volume of the budget** is again represented. In this way the user can ordinarily identify the functionalities of the expense according to their size for each section. |  |
| How has the budget been spent over time? | An area chart t provides an effective and simple graphic representation of how budget is changing using a chronological order. Through this tool, important events in time can be easily seen, compared and understood. A bar chart could also be a suitable representation. Nevertheless, this technique unnecessarily increases the space occupied by the data, in the form of bars, to the extent that users can lose context. | The quantitative variable **total budget** is displayed in the y-axis using an area which fluctuates vertically from the bottom up. If the budget increases, the height of the graph moves further up.<br><br>**Date** is shown in the x-axis as a line that runs horizontally from left to right. |  |
| Where the resources come from? | Bar chart is a good option for comparing the frequency of a large set of categorical variables, in this case the funding sources. The horizontal orientation is useful, so data labels do not look cluttered. We also consider the possibility of using a pie chart but given the number of categories, this does not seem like a clean option. | The categorical variable **funding sources** is displayed as bars in the x-axis. Budget is represented in the x-axis. The lengths of the different bars are proportional to the **total budget** associated to each category. Also, the color of the bars is used to differentiate the volume. A scale ranging from a cold color such as blue to a warm color such as orange was used to indicate incremental budget bands. |  |

| | | | |
|---|---|---|---|
| To which administrative branch does the expense correspond? | A pie chart is a graphic option to represent the numerical proportion of occurrences of a reduced number of categories, in this case the administrative departments expending the budget. Since in this case we only handle three categories then this seems to be a simple, but acceptable alternative to give a clear idea of the proportions. | The **administrative department** is the variable plotted on this graph. Besides the area of the slices, representing the **proportion of the total budget**, color is used to improve the readability of the data. |  |
| Where is the money spent? | A map is used to represent the spatial data given by the provinces, to understand trends or patterns in their expenses. This option is appealing for the users, who are already familiar with the location of interest, since it allows a faster and easy visualization, compared to other valid options as a pie chart or a bar chart. This alternative also allows incorporating elements of contextual interpretation such as cardinal positions or clustering of regions. | The two axes of the graph are latitude and longitude, which constitute the polygons representing each **province**.<br>A third variable is the **total amount of budget** spend in that province, represented by a spectrum of color. A darker color indicates a higher level of spending. |  |
| Who spend the money? | Bar chart is a good option for comparing the frequency of a large set of categorical variables, in this case the institutions spending. The horizontal orientation is useful, so data labels do not look cluttered. | The **corresponding institution** is displayed as bars in the x-axis. Budget is represented in the x-axis. The lengths of the different bars are proportional to the **total budget** associated to each category. Also, the color of the bars is used to differentiate the volume. |  |
| What does each budget component consist of? | A table was selected to display detailed data about the budget items, arranged in columns and rows. In this case, it allowed to show an organized view of the direct relationship between each budget item and other multiple variables such as date, level and status. Graphs are not used since they provide an aggregated view of the data, instead of displaying individual results. | The data fields represent the textual values in rows and columns. |  |

In the following paragraphs we will discuss the decisions corresponding to the interaction between the multiple views.

When manipulating the views, we find a large amount of data (the dataset is very heterogeneous and contains more than 10,000 records). Aggregation plays a key role here given the large amount of existing data. Luckily, the dataset is categorized, so building these aggregations on these categories is straightforward.

The most basic element of interaction that has been used is scrolling. All the elements (using the bootstrap layout) have been placed strategically and in order. The first element that the user finds when accessing the platform is the bubble chart, which is the main component of our visualization. As you scroll, you go through the secondary filters, until the points where the data have the highest level of atomicity: The Bar Chart "Budget by institution" and the final table, where you can see the specific budget item, table that would offer the highest level of detail, and that is affected by all the previous filters.

To reduce complexity, we have implemented filters on aggregate categories, so that the user can intuitively reduce the information to find the answer specific details. These interactive filters also have animations that significantly improve the user experience. In addition, a reduction of them has been used. In the "Budget by institution" chart, for example, the maximum number of institutions shown has been limited. Continuing with the examples, in the Bubble chart, a scale adjustment has been configured, so that semantic zooming is performed when the user applies a filter. By using an animation here, the user maintains a constant traceability of the changes that are applied. Other examples are the use of drag and drop to make cuts in the timeline, which automatically builds a filter between two dates. We have more examples in the Bubble Chart and in the Heatmap, where geometric shapes can be selected.

Finally, it is important to discuss the use of color. In our application we try to follow some best practices like avoiding rainbow color as mentioned in Borland (2007) [2]. We have decided to use a map of a single color with different tonalities to avoid the effect of color blindness.

## 4 USE CASE

In our case of use we will provide the models of insights that can be achieved through our taking advantage of the interactivity of the graphics.

1. How the public budget is spent in terms of purpose?

   As can be appreciated in the bubble chart, the largest component of total spending is social security by far, followed by "Energy, fuels and mining" and "Public Debt Service". Other secondary, but relevant expenses are Transportation, Defense, Farming, and Judicial costs.

2. How has the budget been spent over time?

   To answer this question, we can use the time filters. In our case, we have a line chart that allows us to select time subsets, or use the selector to select a specific year. For this case, we will use the selector, and we will select several views, each with one year: 2015, 2016, 2017, 2018 and 2019. Selecting the debt bubble, and moving to the timeline (Expenses over time), we clearly observe that during 2015, no expenditure was made on public debt. Debt payments were made only between April and May 2016, January and March 2017 and between January and July 2018. This contrasts, for example, with other types of expenses, such as transportation, which is relatively uniform over time.

3. Where the resources come from?

   To find out who paid what, we look at the "Source of financing" bar chart. A priori and by far, the majority spending is supported by the national treasury. However, if we select, for example, social security, we see that it is financed by "Own Resources" and "Internal transfers" mainly. We can also find an interesting fact, and that "Drinking Water and Sewage" is financed mainly with public debt. Spending on Intelligence, however, is financed solely by the national treasury.

4. To which administrative branch does the expense correspond?

   If we want to know which administrative branch makes these payments, we turn to the "Administrative branch" section. If we select, for example, social security, we see how it is mostly paid from "Social Security Institutions", in contrast to the debt, which is paid almost entirely by the Central Administration. Expenses related to energy are paid almost equally between "Decentralized Organizations" and "Central Administration". Spending on Intelligence, for example, is 100

5. Where is the money spent?

   The map "Exepenses by provinces" tells us clearly and directly in which provinces more money is spent. A preliminary view of the map indicates that most of the budget is spent in the capital (state of "Buenos Aires"), while in the rest of the country they have a homogeneous expenditure. If we select this state, we see that the state is spending mainly national treasury resources. An interesting fact that can be observed to the several provinces is that the northern states get more debt than the southern ones, being generally the third source of financing. Thanks to this map, and to the Pie Chart "Administrative branch" it is easy to identify which are the states with the most mineral resources by decentralized organizations, without directly selecting the expenditure in the bubble chart: "Rio Negro" and "Chubut".

6. Who spend the money?

   The following chart "Budget by institution" tells us which specific institution makes the expenditure. For example, selecting "Intelligence" in the bubble chart, we note that it is financed by the Ministry of Security and the Ministry of Defense. Using the same filters we can get some more data: * There is a public institution solely dedicated to public debt. * The Ministry of Defense is the second financier of social security * Legislative expenditure is entirely financed by the National Legislative Power. The Defense Ministry pays part of the transportation expenses.

7. What does each budget component consist of?

   Using a complex filter and looking at the last table, where we can observe the specific budget items (which tell us exactly what was purchased), we can draw detailed conclusions about the expense. For example, there has only been one type of expense ("Telephones, Telex and Telefax") in "Tierra de Fuego", related to mining and energy issues, and it has been in 2017. If we keep playing with the filters, we see that the Higher spending made by the capital, are transfers to other provinces, spending made only between January and March 2017 by the central administration.

## 5 DISCUSSION

In general terms, the visualization has been effective in answering each and every one of the basic questions that can be asked about a budget (how, who, where, when), as well as more elaborate questions that involve the interaction of several variables (As it can be the temporary evolution of the budget destined to a specific function, in a specific province, executed by a specific institution). Therefore, a dataset of thousands of public expense records can be easily scanned. The combination of different technologies (D3.js, DC.js, Python, and MongoDB) not only allowed us to generate a highly interactive platform where the categories of each graph allowed us to filter the budget amount for all the others, it also facilitated the loading and rendering a bulky dataset.

Finally we want to emphasize the fact that designing a project of this kind is not simple. Many factors must be taken into account when proposing a visualization solution. Therefore, during the design phase we try to take into account the phases proposed by Sedlmair (2012) [5] to make the best possible design of views, always keeping in mind the proposed methodological framework consisting of 9 stages: learn, winnow, cast, discover, design, implement, deploy, reflect, and write, as mentioned in the article.

Within the limitations of the application of this visualization solution, we can mention:

- The advantage of the bubble is that it can encode hundreds of data points into a limited area. However, it is less perceptually accurate than other representations like bar charts.
- We consider that our application contains many views on the same page, which could mean an overload for the user. As already mentioned above, our visualization was designed using an Inverted Pyramid approach, which puts the most relevant content up. This could cause that, to the extent that the user moves down, he/she can lose context on the most relevant charts. Therefore it would be important to validate this point by doing experiments with real users.
- Likewise, we are limited to a maximum size of traces that the application can ingest, given the way in which our web server and the database (MongoDB) interact to extract the dataset, as well as the capacity of Javascript and the web browser. For this reason, we do not know how scalable our application is.
- There has been a constant limitation in the amount of charts that we could use, since we started using a library of old dc.js for technical reasons. An attempt was made to update this library but it did not give results, since already implemented methods were used that are deprecated.
- As described by Cockburn (2009) [1], there are many interface schemes that allow users to work and move between centered and contextual views of a data set. Unfortunately we could not add techniques to some of these functions, such as "Approach", which is regularly used to focus areas of the bubble chart where there are relatively small components.

## 6 CONCLUSION

The multidimensionality and complexity of budget data constitute an important challenge in terms of informing citizens and other stakeholders about these issues. Interactive visualization techniques can be an effective communication mechanism so that people without a technical background can understand the structure of the dataset associated with budget items. We consider that our platform offers greater functionality than other tools currently designed for that function as the case shown in the same site where the data were taken: `https://www.presupuestoabierto.gob.ar/sici/quien-gasta`. In this site the visualizations are limited to a sequence of independent treemaps to represent the basic questions about the budget that can only be filtered by the reference year. In this context, the solution discussed throughout this document represents a substantial improvement despite its limitations.

## REFERENCES


[1] A. Cockburn. A review of overview+detail, zooming, and focus+context interfaces. *ACM Computing Surveys (CSUR)*, 41(2):103–119, Aug. 2009. doi: 10.1145/1456650.1456652

[2] R. M. T. I. David Borland. Rainbow color map (still) considered harmful. *IEEE Computer Graphics and Applications*, 27(2):14–17, Jan. 2007. doi: 10.1109/MCG.2007.323435

[3] IMF. Government finance statistics manual 2014, international monetary fund. *World Wide Web*, 1(1):1–200, Jan. 2014.

[4] H. J.-L. . B. D. Jacobs, D. Budget classification. *IMF Technical Notes and Manuals*, 1(1):1–18, Jan. 2009.

[5] M. M. Michael Sedlmair and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Trans. Visualization and Computer Graphics*, 18(12):2431–2440, Aug. 2012. doi: 10.1109/MCG.2007.323435