Visual Analytics: the Bogota Council case

Esteban Dalel R

201227078

John Alexis Guerra-Gomez

February 28, 2018

1   Abstract

Data has been driving the world recently. It is now common to know everything in real time, although institutions in the government is commonly behind. Currently, in Bogota's Council[Castro Gaona, 2016], a group of people are hired to write down everything that happens in single Council session. This trove of information is not actionable, as it is disperse and typed on ancient systems. The files generated are then hand combined and added to a shared repository to be stored[Roncancio, 2012]. Most information in this repository never makes it to any report.

This project will be focused in making the data flow to the correct stakeholders, making the middleman's job easier and dynamizing its use. The forms for data input will be simplified, reports will be automatized, the data will be reorganized and stored for future use, and visualizations will be made regarding the users best interests.

2   General objective

Provide an easy-to-use tool that allows for easy data input and data exploration. This system will allow the user to analyze and compare[Munzner, 2014] data throughout the whole database.

3   Specific objectives

- Make reporting a single day task

    - Through automatized reporting, reduce the workload on the workers assigned to the task

    - Help data recording and annotations by assigned employees to be less prone to errors

- Expand the knowledge available to multiple stakeholders

- Change the way that currently requires a person to navigate deeply the repository to find a single fact and report it to their superiors, by allowing the user to query the data

- Keep repetitive insights front and center to avoid the overhead of asking for it (such as the state of a Alexandria law requirement)

- Categorize knowledge such that non-experts might understand it

  - Summarize the actions required for the usual functioning of the Council, as well as the structure it holds

  - Allow comparison by date, Counselor and Political party of all the political actions that take place

  - Allow users to define their own searches with simple tools that require no technical knowledge (non-SQL)

4   Problem Definition

In Bogota's Council, a colossal amount of information is available. This information is mainly unlinked, disorganized and based upon text that has no clear definition. This makes the work of those in charge of converting it to reports both expensive and time consuming.

All this information is stored in a shared directory hosted in an on-site computer, making data security and integrity a big concern. The tools available to report periodically on the information are limited and, as such, allow only very basic reports such as counts, comparisons, and null marking. There is only one person who knows how to use the Business Analytics tools, so it becomes a bottleneck in the flow of information. The rest of the people involved are in charge of uploading the information to the shared folder by hand, and based upon tables which might not be appropriate for the task.

For data collection there is a XLS file with Visual Basic forms, which generates an unmanageable table that has to be converted by hand into something that the BI tool can handle. Almost everything in the process is prone to errors.

For this project, the team has been commissioned to create an useful reporting tool. This must be both easy to load and to understand. The data flow must not be limited to the few in charge and must give a correct overview of whats happening inside the council, while allowing the user to explore further.

5 State of the Art

Visual analytics has been widely studied, and a lot of use cases arise. In politics, it has been widely use to help decision makers work[Andrienko et al., 2017]. Most of the current work is set on predicting future behavior through Machine Learning, although this is out of the scope of this project.

It has been noted that prior knowledge of the domain does help the user make sense of the visualization at hand. Insight and Creative Thinking are limited to humans[Kijmongkolchai et al., 2017], but expanded by machines, and the two complement each other: machines handle a lot of information and humans make sense of it. As most data is not bi-dimensional, a lot of work has gone to making multidimensional data representation and handling easy and fast[Novellini and Auber, 2007][Sun et al., 2013].

As of now, the best practices[Lu et al., 2017] are related to avoiding user overhead with close to client relationships. When stakeholders are part of the process, their needs can be better understood, and a better product is achieved.

Current solutions in cities are led by teams that believe that connected and open cities are better, such as Sunlight Foundation and GovTrack, which helps teams in US cities such as Boston and Los Angeles give their citizens the data required for overseeing their government. These initiatives put in perspective how helpful for progress is openness. Allowing citizens to be in contact with their representatives, makes them much more accountable.

Similar initiatives are abundant in Bogota, but not visible enough. For instance, Congreso Visible pioneered open data about politic instances, and has been followed by several fact checking and overseeing initiatives. Some examples of private initiatives are Facto, for fact checking, Datos Electorales, for general electoral data,Transparencia por Colombia, an NGO that fights corruption of different kinds and the official Council outlet.

Currently, Bogotas Council is a black box. Even though they do have an information outlet, it is lacking depth, and does not allow for any kind of insights. The proposed project will try, at the very least, to make current and vital information available to the internal stakeholders, while presenting it in a digestible way.

6 Solution Proposal

Develop a system that employs the current data sources to show the stakeholders a simplified version of the data trove they have. This system should be extensible and

replicable under certain parameters.

For data input, a renewed system that reduces the need of navigation and data replication, and allows the lawyers in charge to stop thinking about data input and more of strategy. Cloud first is a must, as instant reports are of the highest importance for the commissioning client.

The system will be deeply user tested amongst as many potential users as possible. People who input the data will be saved from too many fields to fill, and decision makers will have an ever-ready information source that, at a glance, will answer the all too often asked question hows everything?

If the information is deemed public by the owners, the system will be open to the public and will serve as an entryway to Bogotas difficult to understand politics. People will be able to see by zone who are their representatives as well as indicators of what are they doing as a whole. The elected counselors will have a public profile in which their actions will have impact. Political Parties will be visible as how they vote and behave as a group.

7 Calendar

As this project has several external stakeholders, the precise calendar is set weekly, according to the tasks at hand.

- Feb: User interviews, definition of the problem, literature lookup, design thinking with end users

    - Feb 28: Comparison to similar projects, with best practices on the field

- Apr: Analysis of the available info, labeling and summarization, wireframing of the final product

    - Apr 28: All data available will be categorized and its usefulness triaged

- May: Architecture design, programming and light user testing

    - May 28: MVP accompanied of the new input system

- June: Heavy user testing, correction of errors in programming

    - Jun 28: Corrected MVP, documentation

References

[Andrienko et al., 2017] Andrienko, G., Andrienko, N., Chen, W., Maciejewski, R., and Zhao, Y. (2017). Visual Analytics of Mobility and Transportation: State of the Art and Further Research Directions. *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS*, 18(8).

[Castro Gaona, 2016] Castro Gaona, J. S. (2016). *POLÍTICA Y RELACIONES IN-TERGUBERNAMENTALES EN BOGOTÁ, D.C., UNA MIRADA DESDE LA DE-SCENTRALIZACIÓN DE LA CIUDAD.* PhD thesis, Universidad de los Andes.

[Kijmongkolchai et al., 2017] Kijmongkolchai, N., Abdul-Rahman, A., and Chen, M. (2017). Empirically Measuring Soft Knowledge in Visualization. *Computer Graphics Forum*.

[Lu et al., 2017] Lu, Y., Garcia, R., Hansen, B., Gleicher, M., and Maciejewski, R. (2017). The State-of-the-Art in Predictive Visual Analytics. *Computer Graphics Forum*.

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

[Novellini and Auber, 2007] Novellini, N. and Auber, D. (2007). Calcul et Représentation Efficace de Cubes de Données pour une Visualisation Orientée Pixel. *5e Atelier Visualisation et Extraction de Connaissances*, pages 31–42.

[Roncancio, 2012] Roncancio, G. (2012). *Parliamentary Informatics: Building the Foundations for an Analytical Information System about Parliamentary Work using a Data-Warehousing approach.* PhD thesis, Universidad de los Andes.

[Sun et al., 2013] Sun, G. D., Wu, Y. C., Liang, R. H., and Liu, S. X. (2013). A survey of visual analytics techniques and applications: State-of-the-art research and future challenges. In *Journal of Computer Science and Technology*.