

Who dominates the Council of Bogotá and how?

Vivian Aranda, Daniel Ayala, Leonardo Montero

Abstract— The analysis of the behavior presented by the voters is an issue that has gained strength every day, candidates want to understand the results and find information of interest that can lead them to improve their campaigns and detect the areas in which they could have greater influence in future elections. Likewise, several ways of analyzing these data and various visualizations have been created to facilitate the interpretation of said behavior, however, not all of them helps with the analysis and interpretation of the data of the results obtained in the different elections that have been analyzed, that is why it is important to choose correctly the type of visualization and the best channels to be used to be able to find interesting findings. In this project we study the elections to Bogota's Council and made some interest visualizations to analyze it.

For our project we work with data about the electoral behavior of the Bogota citizens, centered in the electoral results to the two last elections for the Council, this data was collected by a citizen's collective of the city of Bogotá called Combo2600 and with information provided by the Registraduría Nacional del Estado Civil (RNEC). We analyzed the data and made some grouping in order to have the same granularity level in all periods of time that we worked with. After that, we made some analysis and elaborate some visualizations that can help to our client to answering some concerns such as Who dominates what localities? How have citizens been vote historically? How has the political map of the city changed? Where is the electorate more open to new candidates and where is it totally dominated by a party or candidate?.

Index Terms— Council of Bogotá, elections, politics, localities, Combo2600.

INTRODUCTION

Combo2600 is a citizen collective that works to promote generational change for the transformation of Bogotá, through the study of issues of the city and disruptive civic actions¹. Thus, it has undertaken an initiative to understand the behavior in the decisions of Bogota voters and also characterize the political map of the city, and as they indicate, understand the distribution of votes of politicians is to understand at the same time their electoral strategies and the interests they will defend once elected. Having a more detailed understanding of that distribution makes it easier to control their management and at the same time identify opportunities to increase the number of supporters in the different localities in future elections.

That is why our interest is focused on being able to support these initiatives in the most appropriate way; with interactive visualizations and analysis of the data presented by the RNEC¹ in terms of results obtained in the voting in the last electoral processes for the Council of Bogotá, we expected to be able to obtain interesting information and present a differential in the information available the candidates to understand their results

Reviewing the literature and the different analyzes and studies that are available on the Internet, not only for elections in Colombia but for elections of different categories in various countries of the world, we have found interesting ways to report the results of the elections and reports that have allowed identify the preferences that voters have and the final results; there are visualizations based on demographic maps, Choropleth maps, Steamgraphs, and histograms that allow to compare the results and electoral tendencies by states or departments.

2. STATE OF ART

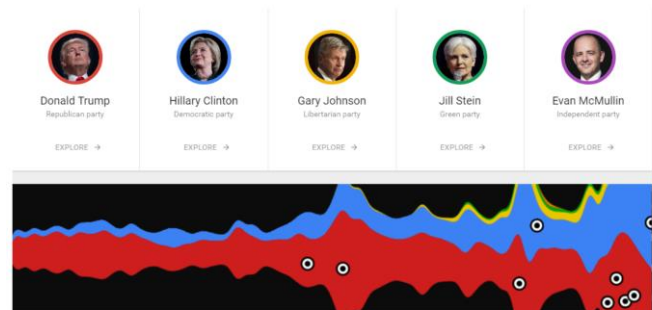
Reviewing pages of newspapers, magazines and political information analysis media we have found several types of visualizations that can be used in our project, as they are adapted to the type of data and information that we want to show.

2.1 Streamgraphs

We find this type of visualizations associated with trends and political choices in the Medium portal. [2]

<https://medium.com/@rcnabho/election-data-visualization-ee041858e7df>.

This idiom allows to represent the information of the votes obtained by each candidate or political party, by locality or for all of Bogotá, and its evolution during each electoral period.



Visualization: Stream Graph

Source: [Google Trends](#)

Technology: JavaScript Libraries

Fig. 1 Streamgraphs.

2.2 Maps

2.2.1 Demographic heatmap

This idiom allows to represent the spatial distribution (by location or for all of Bogotá) of the votes obtained by party or candidate, for each electoral period. We find it as a state visualization associated with elections in the New York Times. [3]

<https://www.nytimes.com/interactive/2014/11/04/upshot/senate-maps.html>

¹ Registraduría Nacional del Estado Civil

North Carolina

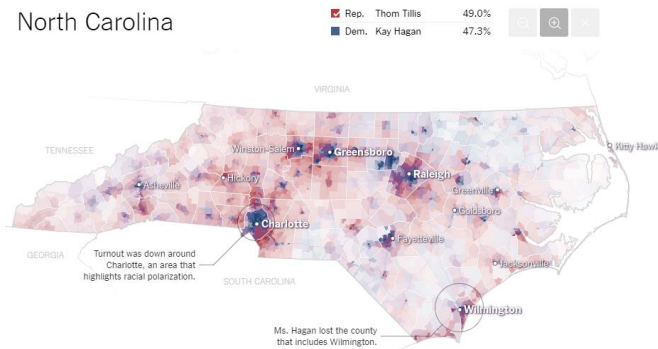


Fig. 2 Demographic heatmap.

This idiom can also be supported by barcharts, tooltips and timesliders to show information of different periods or with greater granularity. <https://codepen.io/sassquad/post/rough-guide-to-building-uk-election-maps-for-d3.js>

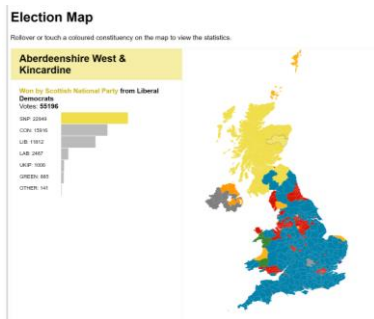


Fig. 3 Demographic heatmap.

2.2.1 Choropleth map

This idiom allows the user to identify geographical areas or regions, making use of colors and intensity. It also allows the user to identify patterns and can be used in particular to identify the population density of voters. In particular, it will identify areas of the city where a candidate or a political party has the most votes.

Choropleth Map

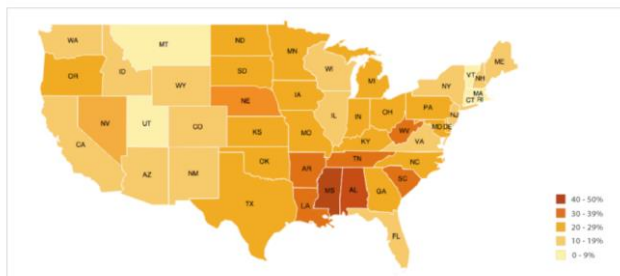


Fig. 4. Choropleth map.

Similarly, in the archives portal of the US government [4] we find another type of demographic heatmap that shows by states the tendency according to the political party. <https://www.archives.gov/federal-register/electoral-college/map/historic.html>

2.3 Parallel coordinates

This idiom is used to plot multivariate numeric data. The graphs of parallel coordinates are ideal to compare many variables and see the relationships between them. It also allows to mix categorical variables with continuous variables and at the same time to establish different scales for the continuous ones. For our case, it can be used to compare the number of votes by locality. 1 <http://bl.ocks.org/syntagmatic/raw/3150059/>.

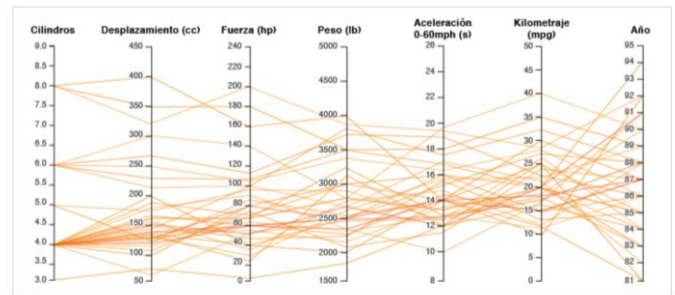


Fig. 5 Parallel coordinates.

https://datavizcatalogue.com/ES/metodos/grafico_de_coordenadas_parallelas.html 1

2.4 Slopegraphs

There are another kinds of visualizations that can help to understand and analyze this data, its slopegraph which are often used to show change over time, in our case it can help to show the quantity on votes obtained by the same candidate in two different periods. We can find examples of its use in politics in some blogs that show tendencies and comparatives along time. 1 <http://guypursey.com/blog/201706282000-slopegraph-uk-general-elections-2015-2017>

3 WHY, WHAT AND HOW

According to Tamara Munzer's characterization in her book Visualization Analysis and Design [2], we have the data abstraction: What, the tasks abstraction: Why and the identification of the possible idioms to be used: How, below we present the results of this initial activity.

3.1 What: Data Abstraction

After review the different sources and datasets we had got at the beginning, we decided to use only the datasets that had been stored in tables, and don't use other external sources, like PDFs reports, because this ones included all information we needed.

Dataset Type: Table

Attributes: (Atributos de interés por fuente)

3.1.1 RNEC – Open Data scrutiny results 2015.

- **Cod_Mun:** Code to the town which the electoral count is made. **Categorical.**
- **Municipio:** Town to which the electoral count is made. **Categorical.**
- **Cod_Partido:** Code of the party to which the candidate belongs. **Categorical.**

- **Nombre Partido:** Name of the party to which the candidate belongs. **Categorical.**
- First and last names of the candidate: **Categorical.**
- **Votación:** Voting obtained by the candidate by town. **Ordered-Quantitative.**

3.1.2 RNEC. Electoral scrutiny results 2011.

- **Partido:** Name of the party to which the candidate belongs. **Categorical.**
- **Votos:** Voting obtained by the party. **Ordered-Quantitative.**
- %: Percentage participation over the total. **Ordered-Quantitative.**
- Potencial sufragantes: **Ordered-Quantitative.**
- Total sufragantes. **Ordered-Quantitative.**
- % Sufragantes: **Ordered-Quantitative.**
- **Partido:** Name of the party to which the candidate belongs. **Categorical.**
- Código: Candidate code. **Categorical**
- **Candidatos:** First and last names of the candidate. **Categorical.**
- **Votos:** Voting obtained by the candidate. **Ordered-Quantitative.**
- %: percentage participation per candidate. **Ordered-Quantitative.**

Storage in separated files for each polling station within each locality.

3.1.3 RNEC - CD with information about elections to council, mayor's office, congress or presidency since 2000

For the years 2011 and 2015 we have received information with a homogeneous structure, with identical attributes and sectorized by zones (sectorization based on the cadastral zoning) which allows comparisons on electoral behaviors.

- **CÓDIGO:** candidate code. **Categorical.**
- **DESCRIPCIÓN:** Candidate name. **Categorical.**
- **ZONA ID:** Identifier for locality
- **VOTOS:** Number of votes obtained by candidate for each zone. **Ordered-Quantitative.**
- **PARTICIPACIÓN:** percentage share of the total of votes **Ordered-Quantitative.**
- **PARTIDOS:** Nombre del partido al cual pertenece el candidato. **Categorical.**

3.2 Why

3.2.1 Actions:

Analyze: Consume: Present → Deployment and visual presentation of available data

Analyze: Consume: Discover → Discovery of electoral behaviors or tendencies not available in existing sources. Formulation of new hypothesis of electoral behavior based on the available data.

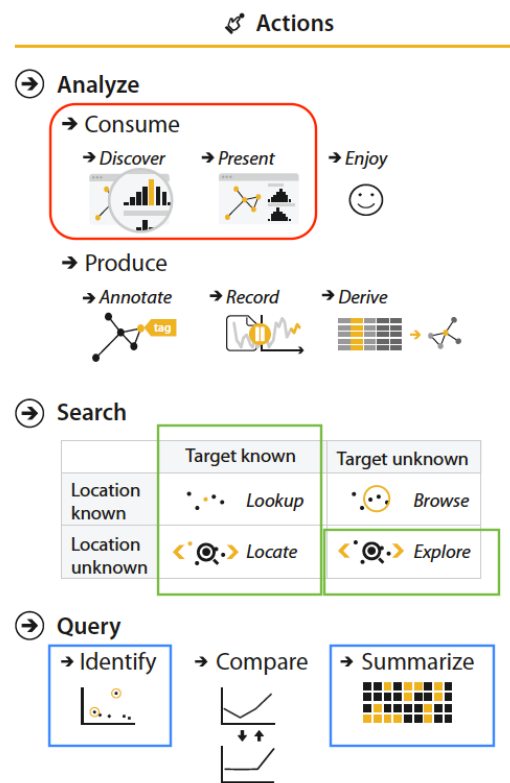


Fig. 6 Actions applicable to the project.

Search: Locate → Find the places (Location unknown) where the most votes have been registered in general or by party (Target Known).

Search: Lookup → Present electoral behavior (Target Known) in certain locations (Location known). Deployment of electoral maps.

Search: Explore → Search of outliers or anomalies in electoral historical behavior.

Query: Identify and Summarize

3.2.2 Targets:

All Data: Trends y outliers.

Attributes → One → Distribution: Deployment of electoral preferences by party

Attributes → Many → Correlation: Correlation between overall results obtained by candidates and their parties by area.

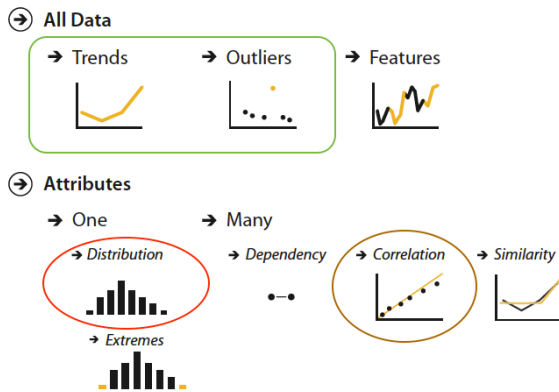


Fig. 7 Targets applicable to the project.

In terms of electoral results of the council of Bogotá we have to be capable to identify some items like that:

- Who dominates which locations?
- How have your votes evolved historically?
- How has the political map of the city changed during las two last elections?
- Where is the electorate more open to new candidates and where is it totally dominated by a party or candidate?

In particular the tasks that we are interested in solving are:

- T1: As a candidate where should I focus my next campaign?
- T2: Identify other candidates similar to me that obtained more votes than me.
- T3: Compare results between candidates from different periods and party.
- T4: Present the difference of results by locality by candidate between 2011 and 2015.
- T5: Locate the area where a political party is strongest.

3.3 How

After reviewing the client's needs and analyzing the data received, we proceeded to choose the data with which we would work and for which periods. It was decided then to work with the information of the periods 2011 and 2015, corresponding to the last two electoral periods; a grouping of the data was carried out in order to obtain the same level of granularity for both periods, given that for the period 2011 the data were in separate files for each polling station within each locality, they were unified to leave them in same terms that the 2015, grouped by location. The final dataset was obtained, which gives us the number of votes obtained by each candidate for each locality, identifying the political party to which each one belongs and the year.

3.3.1 Comparisons of the number of votes obtained by candidate and / or party

Through Parallel coordinates graph we can compare the number of votes obtained sectorized by electoral zone, for each candidate associated to your party for the electoral period selected by the user.



Fig. 8 Paralell Coordinates Visualization

3.3.2 Comparisons of the number of votes obtained by candidate and / or political party

Associated to T1 and T2 tasks using treemap graphs we can compare to other candidates similar to me, in this case we have selected the methodology as follows to select similar candidates: We carried out the segmentation of the candidates where the segmentation variable is the number of normalized votes, the methodology of data mining is k-means, which aims to obtain groups that are homogeneous among them but that when comparing with another group this is heterogeneous. This is achieved by analyzing the distances of the number of votes against the centroid or average of each group, in this way the candidate will be grouped to the centroid to which they have less distance.

The following treemap shows the distribution by sector of the voting of the candidates who by position and percentage of voting, were the most similar in the electoral map to the candidate of the Green Alliance party, Diego Laserna (whose electoral result was 8503 votes) in 2015 campaign.

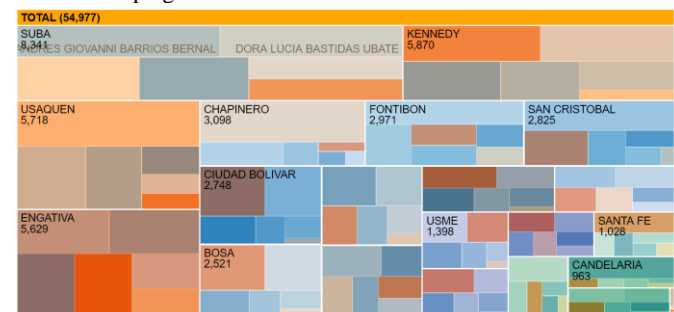


Fig. 9 Treemap Visualization to nearest candidates

In the same way we can analyze the results by locality checking the distribution by political party in 2015.

Comparisons of the number of votes obtained by party in each locality in 2015

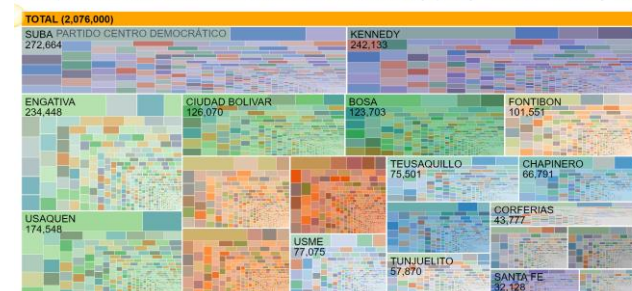


Fig. 9 Treemap Visualization results by locality 2015

3.3.4 Comparisons of election results by period by candidate

Associated to T4 task, using Slopegraphs we can see for each candidate if he/she increment or decrement the quantity of votes obtained along the two different periods. Identifying in which localities improves their performance.



Fig. 10 Slopegraph Visualization results by candidate

3.3.5 Geographic distribution of electoral results :

Associated with task T5, through the Choropleth Map, the need to identify the sectorized behavior of the electoral results is resolved. The client needs to identify the zones or localities in which a particular party has been stronger, in the same way that it allows him to show sectoral preferences for one or another political movement or candidate by the year 2015.

The criteria for the map display are:

- Per year: overall results (all matches, including blank votes)
- By political party: deployment of results by sector

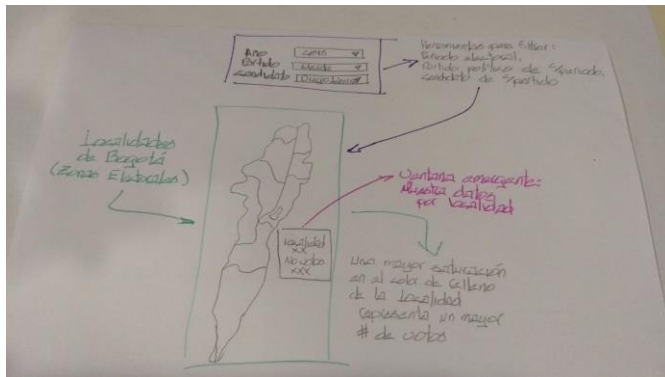


Fig. 11 Mockup Choropleth Map

4 ANALYSIS OF RESULTS AND ANALYSIS BY CUSTOMER

4.1 Data Exploration

About the data exploration and recognition, in order to determine the variables that they contained, the type of variables, granularity of the data for each period, these tasks were performed using Excel, Pandas library (python). Once these tasks were completed, the following preliminary conclusions were obtained:

- Data for all periods allow consulting political parties, candidates elected and not elected by party, total number of votes received per party and candidate.
- The data received for the local elections held in 1997, 2000, 2003, 2000 and 2007 does not allow consulting the number of votes received by candidate and political party for each electoral zone.
- The data received for the local elections held in 2011 allow us to consult the number of votes received by candidate and political party for each electoral zone and for each polling station.
- The data received for the local elections held in 2015 allow us to check the number of votes received by candidate and political party for each electoral zone.

After preliminary review we made tasks of clean and grouping for the 2011 and 2015 periods, in order to obtain the final data structure to performance the different visualizations. We elaborate some *json* files that contains information of candidate including period, their votes, locality, party and personal data.

4.2 Analysis by Customer

As the first interaction of the client with the designed visualizations, we were able to identify some aspects that we could improved.

4.2.1 Visualization 1. Parallel coordinates graph

- It was determined that the selection and selection filter should be improved, it was not so intuitive and the user required an explanation to know how to use it. It also identifies that the use of a mouse from a portable device is not as optimal.
- It should be modified in a way that allows the user to make a comparison between two or more candidates, who do not necessarily belong to the same party.

4.2.2 Visualization 2. Choropleth map

- It was necessary to include a filter by candidate or party
- It is clarified that the type of graph will not allow comparisons but it will be an exploration task that covers this visualization.
- It will be implemented so that it can be visualized as a percentage of the total number of votes received in the city or area, and not only against votes of the same candidate.

As an analysis by the client, it is emphasized that he expects to be able to obtain information about the "electoral barons", identifying if with the passing of the elections they have been strengthened, weakened, have inherited their affinities or have been related with others in order to remain elected. With regard to this, it is determined that the effort should be focused on a certain group of candidates, so that it is for them that this type of atypical behavior can be seen by location and associated with the parties. The ultimate goal is to take advantage of the capabilities offered by visualizations and technology in general to obtain findings from the data and exploit them to the fullest.

Similarly, the alternative is that these visualizations become a monitoring tool that allows to demonstrate atypical behaviors by location.

4.3 Controlled experiment and usability tests

As an academic exercise and in order to make a first validation of the visualizations in which it has been working, a first controlled experiment was carried out in order to validate several aspects of usability and effectiveness of these, based on the tasks, introductory processes were carried out. with the members of the work teams of the course and the visualizations were presented waiting for them to do the exploration of these and they will face the expected tasks and for which they were designed.

As an introduction it was commented that the data had been delivered by an applicant to the council in 2015 and that n had been elected, it was commented that the data had been delivered by the National Registry of Civil Status and that it contained information on the results of different elections, for our project the data of council elections are chosen, in particular those of the years 2011 and 2015, which have levels of granularity that can be worked by area, which is the equivalent to the locality, and which is part of the client's needs. After a small contextualization and with the different users, the following results were obtained for the two visualizations presented.

4.3.1 Choropleth map by locality

This visualization at the time of the test is seen as shown below, it presents a filter by year and another by political party, this with real data for 2015 and simulated data for 2011.

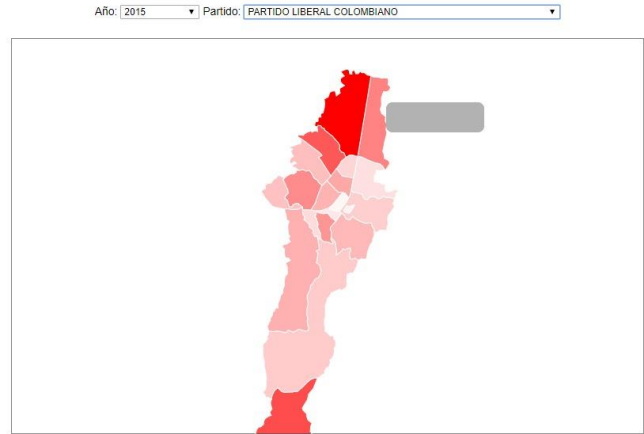


Fig. 11 First versión of Chroloplet map.

- The tasks requested to the different users were achieved by the users, these were related to identifying the area in which a given party had more or less votes in a given period.
- The users make a correct interaction with the visualization, they are interested in the information that appears in the tooltip, they suggest including the percentage of the total in the data shown there, they also suggest to show a table that complements the visualization.
- It is observed that the percentage must be clearly defined as the percentage of the total votes of the locality and not a percentage of the total votes obtained by the party.
- The user validates with what type of information is counted, tries to identify if information is available as a stratum of the table.
- For another user it is not entirely clear the scale of color of the map, it is not easy to see which was the area of least votes of a certain party, likewise it does not understand why the colors, it is clarified that it refers to the political party, example, red for liberal party, green for green alliance, etc. It's funny because other users understood this channel clearly and it seemed appropriate.
- For a user the intensity of the color can represent alarm information, that is, the stronger a very low number of votes can mean, while for the rest it was clear that the higher intensity meant more votes, which was the way to implement this channel.

- They suggest retaining the chosen party by changing the year.

3.4.2 Paralell Coordinates per party per year

This visualization presents at the time of the test a filter per year and another per political party, the source of data contemplates the votes by aspiring councilor.

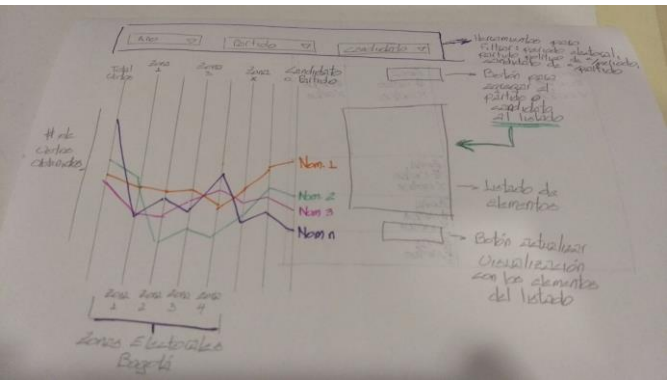


Fig. 12. Primer acercamiento Parallel coordinates.

- The user inquires whether the filters that will be included for additional candidate must be exclusive, it is made clear that they are dependent, that is, once the period is chosen, the parties that had interference in the elections of that period are filtered, and a Once the match is chosen, only the candidates of that party will be there.
- Given that each vertical axis indicates different things, one the candidates and another the scale of votes, it is not clear to the user that corresponds to each axis. Similarly, it is not clear if the scale should be shown for both axes.
- It would be interesting to have two visualizations that allow to see in the top the results by party and in the bottom part the detail until the candidate, not necessarily juxtaposed, but that allows a greater detail without sharing information. This proposal arises for several of the users and indicate that it would facilitate even more the comparison with other candidates and against the party.
- The option to clean the filters must be included, since it is not clear to the user how to do this action.
- The user suggests to perform some type of group by for the different parties, this from a dynamic filter, in order to be able to compare the behavior of a game or candidate against the rest.
- As for the previous visualization, it suggests supporting with an additional table that shows the candidates and allows to make filters from the table and not directly in the graph.
- Suggest lateral selection and not only vertical in the filter.
- The user highlights that there is a correlation between the data presented.

- It is indicated that it is very overloaded, depending on the party, the names of the candidates.
- It is suggested to do some type of filter by clicking on the name of the candidate.

Other general suggestions:

- Match the match logo and the color to the filters.
- Maintain continuity between the different visualizations, which are linked.
- Include appropriate legends and texts that strengthen and facilitate the user's interaction with visualization.

For the final version of our visualizations, we tried to implement all suggestions from the client, the teacher and partners, in general terms we can implemented some of them but all.

5 BINNACLE

Three (3) face-to-face meetings were held with the client, the topics discussed in each of them are detailed below:

5.1. Introductory meeting

This meeting established direct communication with the client, the meeting took place in the surroundings of the University of the Andes, was attended by one of the members of the working group and two people from the collective Combo2600, one of them was Diego Laserna, who was an aspirant to the Bogotá council in the last elections.

The development of this meeting was given by the contextualization of the client to the group in terms of the needs and expected scope of the project, it was specified why it is important for the client to know the areas of the city in which the traditional political parties have greater influence and those that have emerged as alternatives.

A DVD was delivered with data that was generated by the national registry of the civil status in front of the results of the elections to the council and other estates for the years 1997, 2000, 2003, 2000 and 2007. It was also delivered by email of a file in Excel format with complementary data of the local elections for the year 2015, and subsequently a new group of files was submitted in Excel format with data of the local elections for the year 2011.

On the other hand, this meeting identified a clear desire of the client to include map visualizations, as he is aware of the scope that can be given to this type of visualizations and knows how they should be interpreted and exploited.

5.2 Meeting of review of mockups and alternative visualizations

After having carried out a first analysis of the data and the information delivered by the client, a new meeting was held, the objective of which was to review some visualization alternatives that could be useful for the client in the search for their findings.

Although it was indicated to the client that he was looking for each of the proposed visualizations, visualizations already built within the course were used and which allowed to clarify with the client if they would cover the needs of the client. Then showed diagrams of timelines and maps by location of Bogotá in which they were showing counts and trends.

After this meeting what was obtained was the clear definition of what kind of visualizations were valid for the client, being selected visualizations of parallel coordinates graph, choropleth map and slope

graphs to be able to identify variations in the number of votes obtained by each political party and candidate, by electoral zone (if possible), for the different electoral periods from which data were obtained.

5.3 Meeting of revision of first visualizations with data of the client

During this meeting the user was presented with a first approximation of two visualizations that were made from the data delivered by them. It is clear that to achieve this visualization it was necessary to make a large debugging and cross-referencing with other data that were subsequently delivered by the client, it does not have the same level of granularity.

It has been detected that the 2011 data must be derived to obtain the same level of the 2015 data, since these are per table and are required per locality.

The first versions of two visualizations were presented to the client, a parallel coordinates diagram that allows, from one filter per year, to see the behavior of the electoral tendency of the different lobbyists. And another visualization of a Choropleth map of Bogota that allows filtering by candidate identifying the percentage of acceptance that it has in the different localities with marks as the color that indicates the highest percentage of preference.

In this meeting the needs of the client and some expectations that have arisen in front of the visualizations were established and delimited.

5.4 Other interactions and communications with the client

In addition to the various meetings there was telephone communication with the client to clarify some issues of the data that have arisen during the analyzes, and in the same way an email exchange was held, by this means additional information has been delivered. it is used as a source for visualizations.

5.5 Interactions with the teacher

In other validations of the tasks and idioms with the teacher, the observations were received for what was being worked on, for the idea of creating a streamgraph it was received as a suggestion and as an alternative to elaborate a slopegraph that would allow comparing the number of votes obtained by a candidate in 2011 and in 2015.

On the other hand, and after analyzing the data we have and the granularity that we have in them for each available period, the teacher receives an analysis of the tasks, channels and idioms used and, in view of the client's need, it is done an analysis of all of them. It is then proposed to create a visualization that by means of interaction with the user allows to classify the elected candidates of 2015 and based on the characterization of popularity and ideology allows to estimate the votes that a person would obtain if he were a candidate and had an profile similar to classifieds. This application will be available and published under with Google cloud resources.

6 CONCLUSIONS AND INSIGHTS

The analysis of the behavior of voters in the different types of elections is a very interesting task, with the different data sources we can obtain timely results that include characterizations, classifications, analysis of characteristics, and summaries among others, in such a way that we can have results from several perspectives.

Integrating the different sources that we have detected with the sources that will be supplied by the clients of the group resulted in a fairly complete work that provides a complete and quite interesting

picture of how they have captured the intention to vote by political party in the different localities of the city.

The localities with more participation were, in their order, Suba, Engativá and Kennedy. In each of these localities the progressive party and the green party present a very favorable dynamic, thus demonstrating a high acceptance (in proportion to the voters) for the neighborhoods of popular sectors.

For cases like the candidate of the Green Alliance party, Diego Laserna (whose election result was 8503 votes) and the candidates closest to him on the map it can be said that according to the results obtained by these, Laserna was able to strengthen his electoral campaign. In the popular sectors, in places like Suba and Kennedy, the candidate showed very poor results compared to the other candidates. In Usaquén, on the other hand, his participation was important, showing leadership in the counts (vs the other candidates similar to him). His strategy could have been concentrated in the following fronts to increase few votes, 50 more or less, in each of the localities to have obtained the number of votes obtained by the last elected of the green party, Dora Bastidas (9403 votes).

The variety of formats and level of granularity of the data treated leads us to make derivations and groupings of these to achieve homogeneity and improve our analysis, for the period 2015 there was no preferred list for the democratic center which generated a disparity in the analysis of data. In the same way, it was necessary to group the 2011 data at the locality level.

REFERENCES

1. Colectivo Ciudadano Combo2600 página web <http://combo2600.com/>. Recuperado octubre 15 de 2017
2. Registraduría Nacional del Estado Civil, CD con Información de elecciones a concejo, alcaldía, congreso o presidencia desde 2000. Recuperado octubre 15 de 2017
3. Universidad de los Andes (Centro de Estudios sobre Desarrollo Económico - CEDE) Catálogo central de microdatos - Bases de datos Electorales. Recuperado octubre 15 de 2017.
4. Registraduría Nacional del Estado Civil. Catálogo de datos abiertos. Recuperado octubre 15 de 2017
5. Misión de observación electoral. Resultados Electorales En Colombia
6. Universidad del Rosario. Resultados del preconteo para la elección del consejo distrital de Bogotá D.C
7. Registraduría Nacional del Estado Civil. Resultados escrutinios electorales 2011
8. Tamara Munzner. Visualization Analysis and Design. A K Peters Visualization Series, CRC Press, 2014. (<http://www.cs.ubc.ca/~tmm/vadbook/>)
9. Election Data Visualization. Portal Medium. Recuperado octubre 15 de 2016.
10. The Most Detailed Maps You'll See From the Midterm Elections. The New York Times Nov. 6, 2014
11. Registro histórico de votaciones en USA. U.S Electoral College.
12. Nutrient explorer, visualization en blocks.org
13. Portal Dataviz Catalogue, Catalogo sobre visualización de datos, recuperado octubre 31 de 2017.