

1. INTRODUCTION

According to the research carried out by Cuestión Pública in the series "We know what you did", the need arises to show citizens the possible conflicts of interest which may arise when trying to resolve questions such as: Who is the congressman whose family members have the highest contracting with the State? What has been the behavior of the amount contracted over the years? What are the types of contracts with the greatest amount? In which state entities are the amounts contracted and what kind of relationship exists between the congressmen and their relatives? For citizens to be able to resolve these types of concerns, they must have a visualization tool that allows them to know this information, with the goal of being better informed and raising awareness about how they will vote next time.

In virtue of the above and with the participation of an expert linked to Cuestión Pública, different Colombian portals were reviewed in order to determine the state of the art of the available visualizations that address this type of issues, obtaining as a result, the lack of tools that allow to make this type of consultations, for this reason, it was defined as an objective to create a visualization that covers the problematic posed.

To achieve this objective, the following work methodology is proposed and will be described in this paper: (i). Cleaning and understanding of the data, (ii). Definition of users and tasks, (iv). Design and development of an interactive visualization as part of a story telling that allows to explore and discover insights related to the contracting made by relatives of congressmen.

- Cárdenas S. Martha, Student from Universidad de los Andes. E-mail ml.cardenas18@uniandes.edu.co
- Suarez C. Santiago, Student from Universidad de los Andes. E-mail ml.cardenas18@uniandes.edu.co
- Baez C, Journalist and Co-founder of the journalistic portal Cuestión Pública. E-mail claudia.baez@questionpublica.com
- Guerra G. Jhon, Teacher at Universidad de los Andes. E-mail: ja.guerra@uniandes.edu.co

2. STATE OF THE ART

The expert points out that she has not found Colombian visualizations that show something similar to the idea proposed, however, while exploring she found the following:

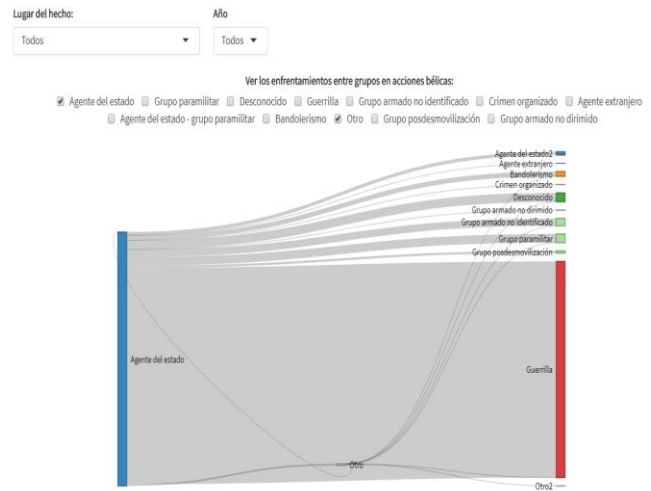


Fig. 2. This visualization is taken from the Observatory of Memory and Conflict [1], which shows the confrontations between groups in armed conflict in the country.

Next, the visualization in figure 2 is analyzed:

Main task:

- Explore the topologies generated by clashes between groups in conflict. (In terms of Tamara Explore - Topology).

Secondary task:

- Summarize the distribution of clashes for each department. (In terms of Tamara Summarize - Distribution).
- Identify the extreme amounts of clashes between two groups (In terms of Tamara Identify - Extremes)

Enjoy:

- Analysis: The visualization shows the relations between the groups in the war actions, for the subject that concerns us, it can be implemented with a better idiom according to the tasks that arise.

3. CHARACTERIZATION

For the characterization process, the framework of Tamara Munzer [2] is used, with the purpose of abstracting to a common language the Why, What and How. Next, the What (abstraction of the data) and the Why (abstraction of the tasks) are described.

3.1. What

Dataset - Network - Trees: Contracting executed with the state by relatives of the ten congressmen selected by the subscribers of the portal <https://questionpublica.com>. (see table 1), which, in turn, correspond to those who obtained the most votes in the last two elections.

Cuestión Pública provided a file of 75 observations with the relation of the amounts contracted by each relative of the selected congressmen, it is worth clarifying that said information was collected through petition rights. Each record presents a link that allows to download a file, which contains the disaggregation of the contracts by family member and contracting entity.

This information was obtained by Datasketch², when consulting the portal Servicio de Contratación Pública (SECOP - www.contratos.gov.co/) using the Scraping technique, which allowed to make a match of the published contracts that had some relation with the relatives of the selected congressmen.

3.1.1. Data Treatment

The following cleaning and standardization activities were carried out on the data set:

- The columns with figures were adjusted to whole numerical values, since some separators of thousands were not interpreted correctly when uploading the data.
- In the numeric columns, the expression (NA, N / A) was replaced by zero.
- Records that contained N / A in most of the columns were discarded.
- The list of contracts was downloaded for each relative of the contractor, in order to relate information to the original data, by contracting entity and contract modality.
- The resulting file of crossing contracts by relative with the list of contracting entities was made up of 292 records

² (<https://www.datasketch.co/>) digital platform dedicated to investigative and data journalism

Attribute	Type
Congresista	Categorical
Votos	Ordered - Quantitative - Sequential
Contratista	Categorical
Años como contratista	Ordered - Ordinal
Vínculo familiar	Categorical
Año contrato	Ordered - Ordinal
Número de contratos	Ordered - Quantitative - Sequential
Valor total contratos	Ordered - Quantitative - Sequential
Entidad	Categorical
Aportó a la campaña	Categorical
Año financiación	Ordered - Ordinal
Monto financiación	Ordered - Quantitative - Sequential
Tipo de contrato	Categorical
Clasificación Objeto Contrato	Categorical

Derive
 Structure - Hierarchy for the tree:
 (Congresista, Contratista (familiar), Entidad) - Categorical
 Agregación Valor total contratos - Ordenado - Cuantitativo - Secuencial

Table 1. Relationship of dataset attributes.

3.1 Why

In the definition it was taken into account that the target user is any citizen who enters the web site of Cuestión Pública in order to find information related to politics and public interest, for which, it is required to offer a tool that allows the citizens to explore and find insights about the way the congressman of their choice has handled their public interest, in relation to state contracting and family ties. Therefore, the following tasks were defined:

Main task (N1): Summarize the distribution of the amount of contracting per congressman made by their families from 2007 to 2018. In terms of Tamara (Summarize the Distribution of variable "Valor total contratos").

Main task (N2): Explore in a fun way the congressmen who have relatives contracting with the state, with their respective characteristics (Amount for contracts, family, company name, number of contracts, types of contracts, etc.) since 2007 to 2018. In terms of Tamara (Explore and enjoy the Features of variable "Congresista").

Task (N3): Compare the contracts by congressman made by their families from 2007 to 2018. In terms of Tamara (Compare the features of variable "Valor total contratos").

Task (N4) Derive family ties between congressmen, contractors and contracting entity. In terms of Tamara (Derive the Topology).

Task (N5): Explore the relationships (family ties) between the congressmen, their relatives and the contracting entities. In terms of Tamara (Identify the Topology of variables "Congresista", "Contratista", "Entidad contratante").

Task (N6): Identify the behavior of the value contracted by the relative of the congressman from 2007 to 2018. In terms of Tamara (Identify then features of variable "Valor total contratos").

Task (N7): Summarize the distribution of the total contracted by the relatives of the congressmen by their hiring modality. In terms of Tamara (Summarize the Distribution of variables "Tipo de contrato", "Congresista" and "Valor total contratos").

Sabemos lo que hiciste (Familiares de congresistas que son contratistas del Estado)

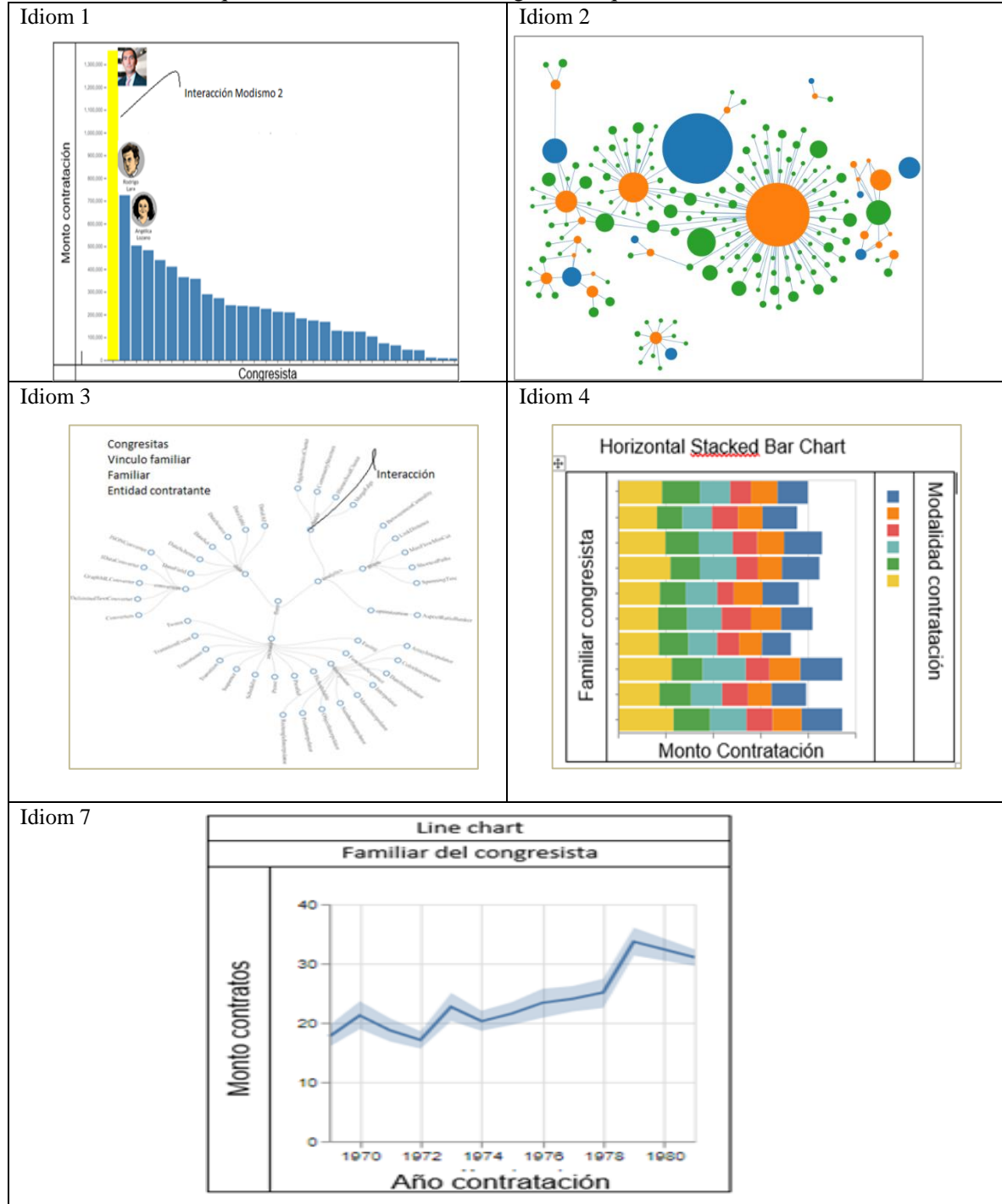


Fig. 3. First Mockup.

Storytelling

ESTOS FUERON ALGUNO DE LOS CONGRESISTAS CON MÁS VOTOS EN LAS PASADAS ELECCIONES



pero....
¿Qué tienen en común estos congresistas?

¡TODOS TIENEN FAMILIARES CONTRATANDO CON EL ESTADO!

Veamos lo siguiente

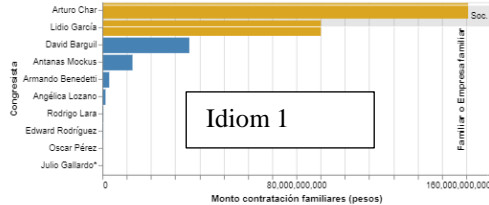
Tip:

Para **explorar** la información del congresista o congresistas, en la gráfica **número uno o dos**, puede dar un **click sostenido** y **desplazar** el mouse sobre la **barra o barras** del o de los congresistas de su elección.

Así mismo, puede observar el **comportamiento** de los montos contratados por los familiares de los congresistas, **seleccionando** el o los **familiares** en la gráfica **número tres**.

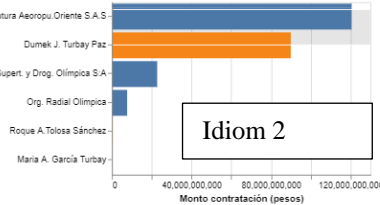
Conflicto de intereses ?

1. Monto Contratado por familiares de los congresistas años 2007 - 2018



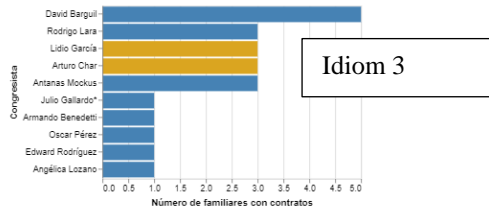
Idiom 1

3. Distribución contratación por familiar



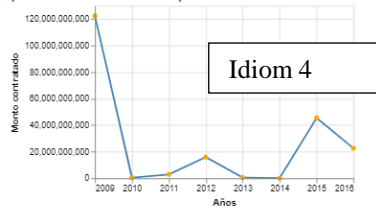
Idiom 2

2. Número de familiares con contratos



Idiom 3

4. Comportamiento de la contratación por año de los familiares seleccionados



Idiom 4



Idiom 5

Fig. 4. Second Mockup: Storytelling that shows insights about contracting done by the relatives of some congressmen.

4. PROPOSED METHOD

According to the what and why described in the previous points, the following proposal was made (see figure 3) of the How (Abstraction of the idioms):

Idiom 1

- Idiom Bar Chart.
 - Mark: Line.
 - Channel: Horizontal position, color (tone).
 - Arrange: Separate, Sort, Align.
 - Attributes: "Valor total contratos de los familiares por Congresista".
 - Description: Represents the distribution of the contracted (family member) value per congressman.
- Applies for tasks N1 and N3 described in point 3.2.

Idiom 2

- Idiom: Radial node-link tree.
- Mark: connection link, point for the nodes.
- Channel: radial axis orientation.
- Attributes: "Congressman", "Contractor (Family member)", "family bond", "contracting entity", "Number of contracts", "Total contract value".
- Description: Represents relationships (family ties) between the congressman, contractor and contracting entity.
- Applies for task N5 described in point 3.2.

Idiom 3

- Idiom: force-directed.
 - Mark: Point for the nodes (Attribute name that contains "Congresista", "Contractor" ("Familiar", "Entidad").
 - Line for the links between the nodes (Source (name), Target (name)).
 - Channel: Encode Arrange Separate and Order.
- | Map | color | Hue | attribute | group, | Size |
|---------------------------|-------|-----|-----------|--------|------|
| nro_contratos_familiares. | | | | | |
- Applies for tasks N5 described in point 3.2.

Idiom 4

- Idiom: Line chart.
 - Mark: Line / point.
 - Channel: Horizontal position, Color (Tone).
 - Arrange: Express.
 - Attributes: "Años como contratista", "Valor total contratos"
 - Description: Represents the behavior of the amount contracted by the relative of the congressman over time.
- Applies for tasks N6 described in point 3.2.

Idiom 5

- Idiom: Stacked bar chart.
- Mark: Line.
- Channel: Horizontal position, Length, Color (tone).
- Arrange: Separate, sort attributes: "Contratista" ("Familiar congresista"), "Tipo contrato", "Valor total contratos".
- Description: Represents the total amount contracted by the relative of the congressman, as distributed by the type of

contract.

Applies to task N6 described in point 3.2.

5. RESULTS AND EVALUATION

In relation to the proposed idioms (see figure 4), the following comments were received from the expert:

Idiom 1

She agrees that the idiom allows identifying the congressman who has the highest amount of contracting done by his family and suggests adding the attribute of the number of relatives who have entered into contracts with the state. On the other hand, she suggests that it should be the first idiom that appears in the Storytelling.

Idiom 2

The use of circles is not the most appropriate for the user, since when evaluating the clicks of the Web site of Cuestión Pública, people do not navigate on that type of visualization. Also, it indicates that the visualization is not very clear for users.

Idiom 4

It displays a timeline with the amount of recruitment for each of the family members of the congressman, however, the expert indicates that it is necessary to include (add to the data) the year in which each congressman was chosen, to see if the relative began to contract from that same year or later, likewise, if the amount contracted has increased.

Idiom 5

The expert indicates that the modality of contracting by itself, for the type of user, does not say much, for which it is necessary to add the typification of the contract object to the data.

According to the defined idioms and the observations received, the first version (figure 4) of the visualization was created with the incorporation of Storytelling. With this version, usability tests were carried out with Visual Analytics students and the client. Main results are detailed below.

Storytelling

- Users suggest incorporating the name of the congressman in each photo.
- They like the presentation and the color of the slides, but it is suggested that when they do not interact, the time to observe the slides is very short.

Idiom 1

- Users have difficulty comparing the contract amounts per congressman, since full bars are not shown.
- Users state that an indication is necessary to interact with the visualization.
- The users identified which one is the congressman who presents the largest amount of contracts by their family members.

Idiom 2

- The users identified which family or family company has the highest amount contracted.
- Users are confused with the order of the graphics, suggest reorganizing them.

Idiom 3

- Users can identify the congressman who has more relatives in contracts.
- It is suggested to change the order of the graphics by the interaction that is carried out.

Idiom 4

- Users find it interesting to look at the relatives and the amount contracted over the years.
- They suggest revising the scale because it confuses when it displays the year repeatedly.
- They suggest changing the order for the interaction.

Idiom 5

- The network thinks this is an interesting way to visualize the framework between the congressman and family and indicates that it would be important to show the type of family bond, likewise the companies or entities with which they contracted.
- Add tooltip in the network that summarizes more details and totals of the information of each congressman.
- The user indicates that when selecting several nodes, some nodes are lost from view.

COMMENTS FROM THE EXPERT

- It is necessary to incorporate a guide for the interaction with the visualization.
- In relation to the scale, she suggests using thousands of pesos instead of pesos for more understanding.
- Indicates the need to incorporate a tooltip that provides more information to the user.
- Suggests revising the scale in idiom 3.
- In the two graphs that appear, she advises to leave the first one visible when the interaction begins, to avoid confusing the user.
- She recommends adding the amount of years as a congressman in the tooltip of the idiom 3.
- Suggests replacing the photos with Cuestión Pública illustrations in the Storytelling.
- Advises to incorporate the graphic identity of the client.

Figure 5 summarizes the evaluation of each task and shows that tasks N5 and N6 defined in point 3.1

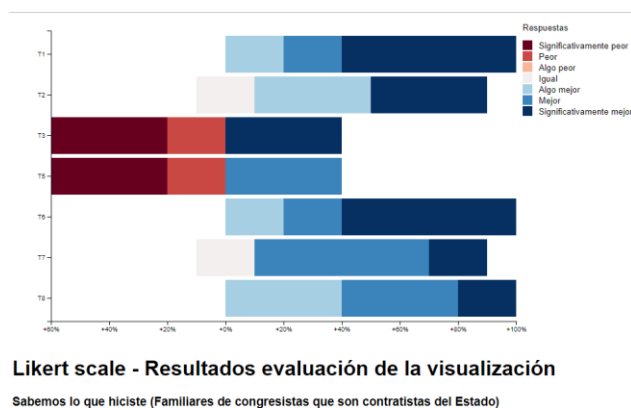


Fig. 5. Results from evaluation of the visualization.

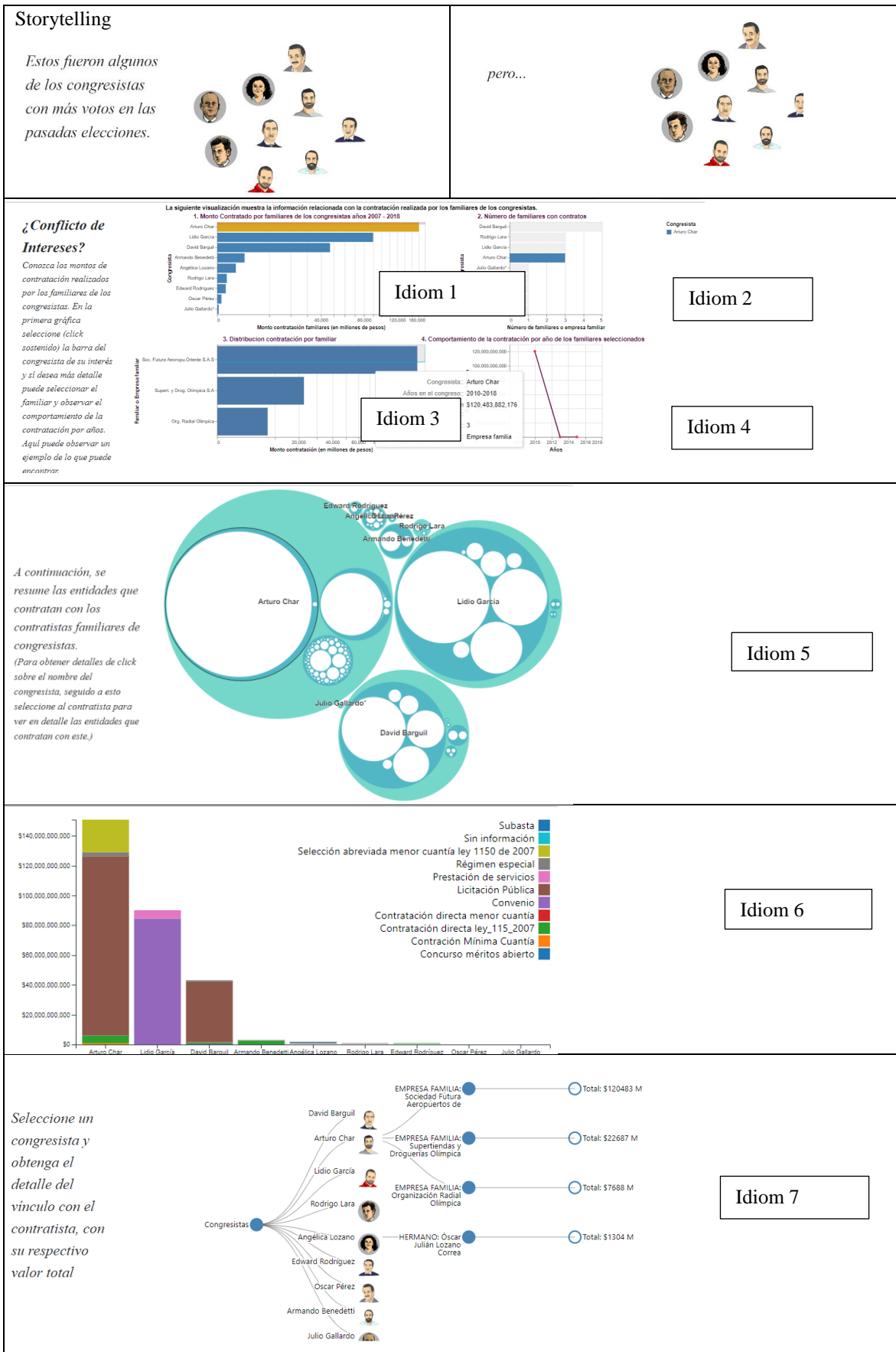


Fig. 6. Final version: Storytelling that shows insights about contracting done by the relatives of some congressmen.

6. CONCLUSIONS AND DISCUSSION

Once the changes have been made as a result of the usability tests, a final version is generated (Figure 6). Here are the adjustments made:

- Because coordinated views are used, the order of idioms was adjusted
- The color of idiom number 2 was changed.
- The scales of the idioms were adjusted.
- The tooltip was added with more information from the congressman.
- It was difficult to explore relationships in the idiom of the network; therefore the park circle idiom was created, which allows to see the relations between the congressman, his or her relatives and the entities which they made contracts with, in an interactive way.

6.1 CONCLUSIONS

When analyzing the implementation of the six idioms in storytelling, the following significant insights were identified:

- The congressman linked to the largest amount of contracts due to his family or family business is Arturo Char, with a value of \$ 150,000 million pesos in 416 contracts.
- The Congressman who has the largest number of family members with State contracts is David Barguil, with 5 family members and 32 contracts with a contract amount of \$ 43,000 million pesos.
- The family business "Sociedad Futura Aeropuertos de Oriente S.A.S" of congressman Arturo Char (congressional periods 2010 - 2018) has generated the largest amount contracted, amounting to more than \$ 120,000 million pesos in three contracts; hiring started in 2010 and the highest amount was received in the same year.
- The brother of Congresswoman Angelica Lozano (congress periods 2014 - 2018) presents a contract that began in 2015 for 185 million pesos and amounts to 987 million pesos in 2017, for a total of 1,304 million pesos for 13 contracts.
- Congressman Arturo Char presents the highest concentration of entities with which the family company entered into contracts. Followed by Congressman Lido García and David Barguil.
- The type of contract with the highest contracting made by the relatives of the congressmen is the Public Tender, followed by the Agreement.

REFERENCES

- [1] C National Historical Memory and Observatory of Memory and Conflict. See the clashes between groups in war actions. <http://centrodememoriahistorica.gov.co/observatorio>. 2018
- [2] T. Munzner. Visualization Analysis and Design. Taylor and Francis Group, 6000 Broken Sound Parkway NW, SUITE 300, 1st ed., 2014. doi: ISBN 978-1-4665-0891-0