Fulbright Foundation Scholarship Evolution

Mera D. David, Polo P. Arturo, and Quintero C. Omar

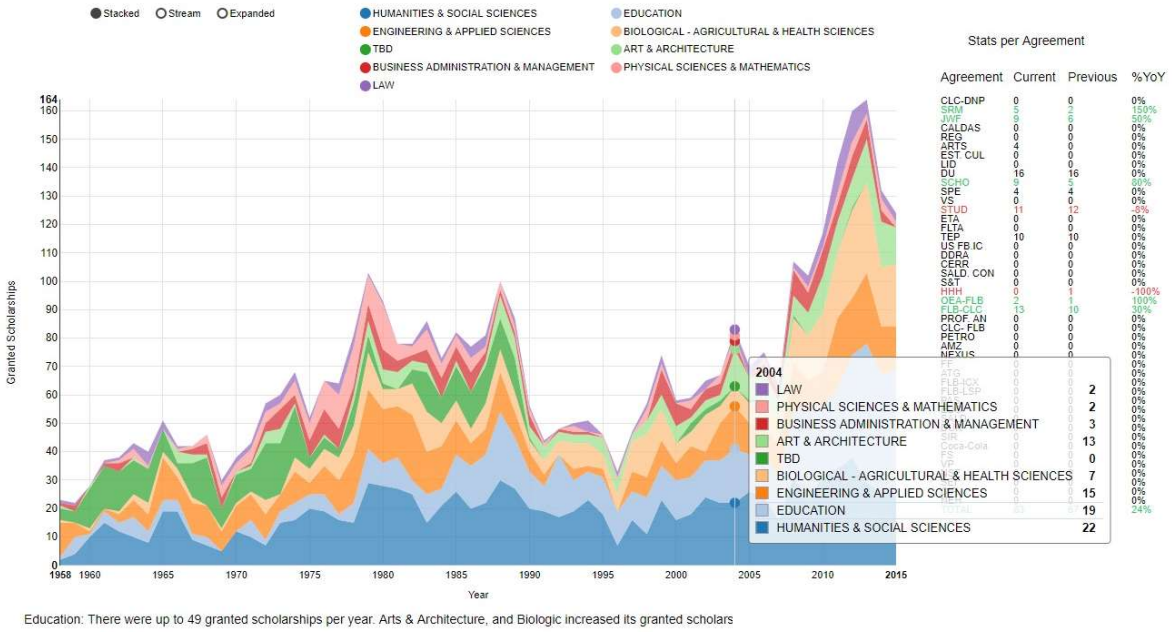

Fig. 1. Initial view of visualization

Abstract— The interest of the FULBRIGHT Foundation to analysis the data they already have is to evidence the evolution and improvement in the number and distrbution of the scholarships granted ever since its inception with the objective of identify which attributes are the most signifcant and through this way validate the current strategie focused on reach more and better candidates and increase the effectiveness percentage of the applications. After the analysis of the requeriments, three different visualiations were proposed: Multilines Chart, Parallel coordinates and Radial Netword Diagram. The Multilines Chart would present the evolution of the scholarships, the Parallel coordinates chart is suitable to identify the significant attributes and the Ring Network Diagram allows us to find the relationships between Universities. In this way we expect to fill and cover the Fulbright Foundation needs of insights.

Index Terms—Sholarships, University, Fulbright, Colombia, Visualization, Visual Analytics.

INTRODUCTION

The FULBRIGHT scholarship program has promoted the knowledge exchange between USA and the rest of the World since 1946. In Colombia, since 1957, many Colombians and Americans have participated yearly in academic and professional activities who has contributed to a better understanding between two countries.

With the target of develop the all-inclusive access politic and the searching of new sources of founding, the analysis of the data who already has the foundation takes on a huge importance in order to be able to identify and/or anticipate the courses of actions.

In this document, we explain the Visual Analytics application we built for the FULBRIGHT Foundation that brings some insights to increase the current knowledge of his current context and allows them to take better decisions to reach the goals they have.

1 TASKS ABSTRACTIONS

1.1 Main Task

To show the evolution of the scholarships granted by the FULBRIGHT Foundation since 1958 to identify the growing of these based on the new management and the subscriptions of new agreements**.**

1.2 Secondary tasks

- ***To know*** the distribution and diversity of the socio-demographic features of the FULBRIGHT foundation scholarships grantee taking into account aspects like its university, educational level, gender and geographical region where he comes from.
- ***To find*** the strength of the relationships between Colombian and Americans universities in order to suggest which should be the better strategy for each agreement.

2 WHAT

2.1 Dataset

Table with 4.002 records belonging to ex grantees of a FULBRIGHT foundation scholarship. Some attributes have been encrypted to allows its manipulation guarantying the accomplish all the confidentiality requirements of the foundation.

2.2 Attributes

The original dataset had 82 attributes. As part of the applied methodology, a strict cleaning process was done and 44 attributes was excluded:

- 36 attributes with a completeness less than 10%.

- 8 checking fields sent by the Foundation.

Finally, only 38 attributes were taking account inside the analysis done. These attributes have the next characteristics:

Attribute	Type	Completeness
Cohort	Sequential	100%
Conv.	Sequential	100%
Fulbright Grant	Categorical	100%
Abbrev	Categorical	100%
Id	Categorical	100%
Date Birth	Sequential	12%
Degree	Categorical	100%
General Academic Area	Categorical	92%
Field of Study	Categorical	92%
Colombia Grantees – US University	Categorical	65%
Subfield Study/Degree Program Name	Categorical	43%
Colombia Grantees – State	Categorical	64%
Colombia Grantees – City	Categorical	64%
UD Grantees – Host Institution	Categorical	24%
UD Grantees – Department	Categorical	25%
UD Grantees – City	Categorical	25%
Gender	Categorical	99%
Country of Nationality	Categorical	100%
City of Residence	Categorical	56%
State of Residence	Categorical	56%
City of Birth	Categorical	25%
State of Birth	Categorical	26%
Degree	Categorical	73%
Field of Study	Categorical	76%
University 1	Categorical	76%
Type (Public/Private)	Categorical	43%
Degree	Categorical	23%
Field of Study	Categorical	24%
University 2	Categorical	24%
Type (Public/Private)	Categorical	10%
Highest Program Degree	Categorical	73%
Program Name	Categorical	11%
General GPA	Sequential	11%
Name of Employer	Categorical	22%
Sponsoring Institution	Categorical	19%
Academic Program Period Start	Sequential	14%
Academic Program Period End	Sequential	14%
Status	Categorical	89%

Table 1. Working dataset

2.3 Availability

The dataset is static and is entirely managed by the FULBRIGHT foundation. The applied schema allows the foundation to update the dataset every time they want.

3 WHY

- ***Present*** the evolution of the scholarships given by the FULBRIGTH foundation to Colombians since 1958 with the objective of evidence the different trends over the years.
- ***Present/Compare*** the distribution of the educational and demographic aspects of the FULBRIGHT foundation scholarships grantees in order to demonstrate the diversity of the candidates over the years.
- ***Identify*** the existing relationships between Colombians and Americans universities based on the number of scholarships granted by the FULBRIGHT foundation.

4 HOW

4.1 Stacked Area Chart

To support the main task, we propose a STACKED AREA CHART together with a table. This idiom allows the user to identify the numbers of scholarships granted by the Foundation in a specific period and its evolution/variation over the years.

Marks:

- Stacked areas.
- Glyph: composite object, internal structure from multiple marks.

Channels:

- MAGNITUDE
  - Horizontal position (ordered and separated years).
  - Vertical position: number of scholarships granted.
- IDENTITY
  - Tone for each one of the areas of study.

Manipulate:

- CHANGE → REALIGN: the idiom proposed lets the user the possibility to realign the visualization in three different ways: Stacked, Stream or Expanded. This feature was implemented to make the identification of trends or patterns easier to the user.

Facet:

- JUXTAPOSE → MULTIFORM, OVERVIEW/DETAIL: both the chart, the table and the popup works together to present the user more information about his navigation. The idea of this design was let the user to browse the stacked area chart to identify and specific period and include detailed information through the popup, regarding with the categories, and the table, regarding with the agreements.

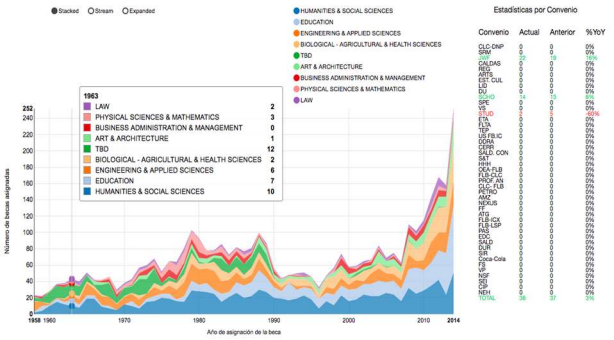

Fig. 2. Stacked Area Chart proposed.

4.2 Parallel Coordinates

To support the tasks of present and/or compare the demographic aspects of the FULBRIHT scholarships grantees we propose a Parallel Coordinates diagram. The idea is to have one vertical line per dimension and one horizontal line per scholarship granted. The idiom proposed will allow the dynamic filtered of the relationships displayed.

Marks:

- Vertical parallel lines as axis representing dimensions
- Point to represent each item (scholarship granted).

Channels:

- MAGNITUDE
  - Vertical position to represent the number of FULBRIGHT foundation grantees.
  - Horizontal position that represents the value of the dimension.
- IDENTITY:
  - Tone to identify the scholarship

Manipulate:

- SELECT: the user can select and highlighted one or more horizontal line to emphasizing its importance.

Facet:

- JUXTAPOSE → MULTIFORM, OVERVIEW/DETAIL: both the parallel coordinates chart and the table works together to present the user more information about his filtering action. Once the user applies a filter over a dimension, an event that update the other dimensions of the

parallel coordinates diagram and the detailed table is triggered.

Reduce:

- FILTER: the idiom proposed allows the application of a BRUSH over a dimension filtering its values and showing only a portion of its records.

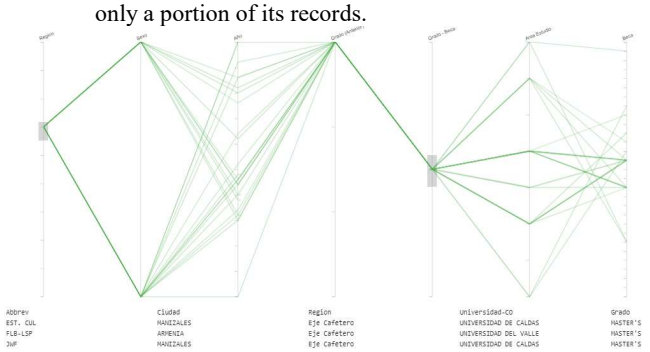

Fig. 3. Parallel Coordinates proposed.

4.3 Ring Network Chart

To support the task of identify relationships between universities of Colombia and American universities we proposed this visualization. The idea is to uses arrows between two or more nodes (universities) to represent the relationship.

Fig. 4. Ring Network Chart proposed.

Some additional features were included in this implementation to improve the user experience. These features are:

- Filter by year
- Filter by university source
- Filter by university Destiny
- Change variables (include or exclude)

Figure 3, shows links in a specific year for Eafit university. In this example, it is easily evident that the EAFIT university made five (5) agreements with a foreign university.

Marks:

- Circles joint by an arrow

Channels:

- IDENTITY: Tonality for the different universities

Manipulate:

- SELECT: through the circles you can select a university

Facet:

- JUXTAPOSE → MULTIFORM, OVERVIEW/DETAIL: through the RING NETWORK a first glance is made of the relationship that exists between universities that through Fulbright have exchanged students, additional menu allows filtering on a specific university what triggers the updating of the graph.

Reduce:

- FILTER: allows the user to select which columns to display as well as the order in which they are displayed. The user may also filter on specific column values; both categorical and numeric

Fig. 5. Filter Ring Network Chart proposed.

5 State of Art

5.1 Discussion of Multivariate Analysis

The most popular techniques that can be implemented to present and/or do multivariate analysis are:

5.2 Data Tables

This is simplest way of all. Each row is an occurrence of the data, each column a variable and each cell (intersection between row and column) reflects a value of that variable for an occurrence.
Tables are ideal to present values but is not the most efficient way to identify pattern, groups, correlations or trends between variables.

5.3 Scatterplot

This chart consists of two axes that represent each of the variables. The intersection value between these two variables constitute a mark. There is a possibility to extend this visualization to multiple variables through the construction of a matrix. The idea is to distribute the dimensions in the vertical and horizontal axes and perform a scatterplot for each of the intersection.

The scatterplots work correctly if the objective is to locate the correlation between variables based on the shape configured by the points. In the same way, the task of detecting outliers, detecting patterns are include in its strengths.

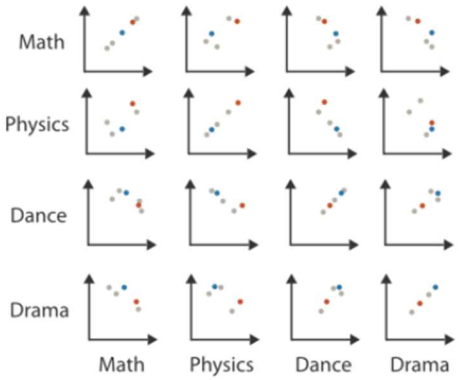

Fig. 6. Matrix Scatterplots

5.4 Parallel Coordinates

Consists of two or more axes aligned vertically or horizontally, which represents the dimensions to be included in the analysis. Line segments that intersect two axes represent data occurrences.
This type of visualization allows the possibility of performing continuous and comparative analyses on N variables represented as axes with the ability to identify the different groups, correlations or trends among all variables.

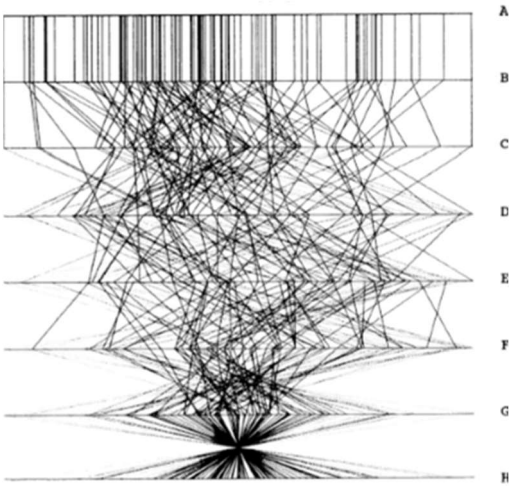

Fig. 7. Parallel Coordinate

If the line segments between two axes are parallel, it means that there is a positive correlation between these two variables. If the lines segments cross each other in the middle point of the axis, it means there is a negative correlation between these two variables. If there is no a clear pattern in the line segments between two axes, there is no correlation defined between these two variables. In the figure five (5), the correlation between variables A and B is totally positive, while the correlation between the variable G and H is totally negative.

5.5 Tree Map

This visualization uses a hierarchical characteristic of the information to present it. Size, annotation and location of the nodes play an essential role for the identification of patterns. This type of visualization works if the data has a hierarchical structure. It is an appropriate visualization when the most important is the size or weight of the variables.

6 Implementation

Given the characteristics of the requirement to be performed and the number of variables to be included, it is concluded that:

- A data table is not viable since it does not allow the identification of correlations, patterns, grouping and outliers.
- A Tree Map is not viable since, despite being optimal for the visualization of clusters, it does not work for the identification of correlations and we do not have data that have a completely hierarchical structure.
- A Scatterplot works for all the requirements we have; however, we should consider using a Scatterplot matrix taking into account the number of variables.
- The visualization of parallel coordinates works also for all the requirements that we have and offers the possibility of the inclusion of multiple variables.

According to Henley in [1], the Scatterplot diagram is better than the parallel coordinate diagrams, but they have the limitation of only focusing on bi-varied analyses. Extrapolating the solution to a Scatterplot matrix adds complications with the presentation and easy identification of the characteristics.
According to Kuang in [2], parallel coordinate diagrams have an advantage over Scatterplot when the dataset to work has a low number of records (low density) and the level of dimensions is controlled.
Taking account that both the parallel coordinate diagram and the Scatterplot comply fully with the stated requirements, we choose to use the first instead of the second for the following reasons:

- Facility to include multiple variables within the same diagram, which contributes to the better understanding for the end user.
- Information scale.
- Number of known and mapped dimensions.

The visualization planned locates the dimensions as vertical lines parallel to each other and represents the occurrences, that is, the scholarships delivered by the FULBRIGHT foundation since 1957, as lines that intersect these vertical axes.
Some authors such as Theisel in [3] mention the possibility of using curves instead of line segments to connect the vertical axes based on for a person it is easier to visualize or follow curves than lines. However, the complication of the identification of correlations emerges, which is much more intuitive using lines in the way previously described.
The implementation would use a visualization of parallel coordinates as an idiom, line segments as marks and tone, horizontal and vertical position as channels.

7 BLOG

Date: 2017/09/13
Meeting participants:

- Greis Cifuentes – Community Coordinator.

Details: The first contact is established to schedule the meeting with the Foundation officials who are interested in making the project.

Date: 2017/09/22
FULBRIGHT-UNIANDES
Meeting participants:

- Natalia Prieto – Community Coordinator
- Felipe Zambrano – New Projects Coordinator
- Adriana Prieto – Program Director

Details:

- FULBRIGHT coordination explains all about the business questions they would like to solve trough this project.
- Tasks for the visualizations were identified.
- The commitments of each of the parties were delimited.

Date: 2017/09/27
FULBRIGHT - UNIANDES

Meeting participants:
- Natalia Prieto – Community Coordinator

Details: Agreements were made regarding the confidential use of information provided by the FULBRIGHT Foundation. The database of ex-grantees was also obtained, and a first approach of the visualization was made through mock-ups.

Date: 2017/10/01
Group Meeting
Details:
- It is carried out in order to clarify doubts and socialize conclusions of the meeting held with client on 27/09.
- The tasks are distributed to each member of the group

Date: 2017/10/15
Group Meeting
Details:
- Sample of advances and internal feedback
- Adaptation of Ring Network Chart.

Date: 2017/10/30
Meeting participants:
- Natalia Prieto – Community Coordinator
- Felipe Zambrano - New Projects Coordinator

Details:
- Mock-ups developed with real data of the client were evaluated using Google forms.
- Users make general comments about the project's objectives and visualizations to adapt them

Date: 2017/11/07
Group Meeting
Details: Definition of internal visualization tasks and idioms in the group. Responsibilities and functions.

Date: 2017/11/25
Group Meeting
Details: Internal follow-up meeting.

8 EVALUATION AND FINAL USERS CONSIDERATION

The way to evaluate the solution proposed was chosen based on the skills and experience of target population. The people from the FULBRIGHT Foundation has a minimal knowledge about visualization techniques application.
A first and controlled experiment was developed, exposing the client to visualizations and let them explore and makes their own conclusions. After this step, each one was asked about the result of his interaction, if they consider these were useful or not and what improvements should be implemented.
The survey applied had the following questions:
- Regarding your experience manipulating the visualizations, how did you feel?
- Regarding the same graph, what was your feeling about manipulation of it?
- Would you propose any improvement?

STACKED AREA CHART:
- It was agreed that it allows you to see the evolution of each of the FULBRIGHT scholarships and is extremely practical for the purpose they are looking for.
- As a recommendation, it suggests that the inflection points of each year show the number of fellows of each cohort

RING NETWORK:
- Super easy to use, extremely visual, comfortable to analyze.
- Reports are achieved with a click ... What can I say?
- Yes, the consolidation of some years in which these relationships are shown is still lacking, and it would be even better if the user could define the period and it was not predefined by tabs.

PARALLEL COORDINATES:
- Users expressed that their handling is pleasant and useful
- When manipulating it, they felt it was simple and clear.
- As a recommendation, they requested dynamic update of variables

9 INSIGHTS

As insights, we would to highlight the next ideas:

First, the strategy methodology applied by the Foundation over the years could be splitted by decades:
- Before the 80's, most of the data related with field or subfield of study were not captured, thereby it's very complex try to identify which was the key factor of the behavior registered during this period of time.
- In the 80's, the focus was the Human and Social Sciences scholarships. 25 scholarships average per year were granted in this field of study. This scenario was driven by the increase of the scholarships granted of some agreements called ATG (AIRWAYS TRAVEL GRANT +600%), FLB-LSP (BECA FULBRIGHT-LASPAU +220%), SIR (BECA PROFESOR COLOMBIANO EN RESIDENCIA +1400%) and CIP (COUNCIL OF INTERNATIONAL PROGRAMS +1000%).
- In the 90's, the focus was to reduce Engineering scholarships. Between 1990 and 1996, the engineering scholarships suffered reduction by 73%. The J. WILLIAM FULBRIGHT agreement, who had granted 29 average scholarships per year in the past, during this period only contributed with 5 scholarships average.
- In the latest 90' and till now, the focus has been the education scholarships. The new boom is the agreement called BECA ASISTENTE EN LA ENSEÑANZA DEL INGLÉS. Just through this agreement from 2008 till 2015 were granted same amount of scholarships (150) than the number of scholarships granted by all the agreements between 1997 and 2007.

In other plane, we could find the next insights:
- Since 1980 the number of women has increased. However, none of them had Ph.D. as a previous degree. Also, they haven't applied for scholarships in Arts and Physical Sciences.
- Since 1987 alumnus from Amazonia region has been interested in scholarships related with Health, Biologic, Humanities, and Social Sciences. Nonetheless, the last woman from Amazonia who applied for a scholarship was in 1997.
- Since 2010 the number of granted scholarships in Pacific Coast and Valle de Aburra regions has been increased.
- In Pacific Coast region nobody has applied for scholarships in Arts and Business Management.
- In Valle de Aburra region nobody has applied for scholarships in Humanities and Social Sciences.
- The University of Antioquia (1803), despite being one of the most important in Colombia, during the first years shown less agreements than Universidad Industrial de Santander (1948).

10 CONCLUTIONS

As conclusions, we would like to mention the major achievements that can be summarized:

- The architecture was designed to work 100% from the browser and developed in D3 (JavaScript). Some external libraries were included into the Project like NVD3 to enrich the visualization features that was offered to the user.
- At the beginning the idea was to enable the interaction with the FULBRIGHT foundation through a Google Doc file but due to reasons of privacy policies, it was impossible. Finally, we took the decision to avoid this solution and work locally with the data. That means the FULBRIGHT foundation
- To accomplish all the requirements, we choose some types of diagrams. Nevertheless, we had to enrich these diagrams with other features like tables, filter, selector, labels, etc. to solve the business questions the foundation had.

REFERENCES

[1] M. Henley, M. Hagen, and R. D. Bergeron. Evaluating two visualization techniques for genome comparison. In Proceedings of 11th International Conference Information Visualization (IV ’07), pages 551–558. IEEE, 2007.

[2] X. Kuang, H. Zhang, S. Zhao, and M. J. McGu n. Tracing tuples across dimensions: A comparison of scatterplots and parallel coordinate plots. Computer Graphics Forum, 31(3):1365–1374, June 2012.

[3] H. Theisel. Higher order parallel coordinates. In Proceedings of VMV, pages 415–420, 2000.

[4] R. Kanjanabose. An Empirical Study on Parallel Coordinates and Scatter Plots. A thesis submitted for the degree of Master of Science in Computer Science. University of Oxford. Pages September 2014.

[5] J. Heinrich and D. Weiskopf. State of the Art of Parallel Coordinates. In Eurographics 2013 - State of the Art Reports, 2013.

[6] J. Brickell and V. Shmatikov. The cost of privacy. Knowledge discovery and data mining, 2008

[7] A. Dasgupta and R. Kosara. Adaptive Privacy-Preserving Visualization Using Parallel Coordinates. Transactions on Visualization and Computer Graphics (Proceedings InfoVis), vol. 17, no. 12, pp. 2241–2248, 2011

[8] [26] L. Sweeney. k-Anonymity: A Model for Protecting Privacy. IEEE Security And Privacy, 10(5):1–14, 2002.

[9] [27] S. F. V. Ciriani, S. De Capitani di Vimercati and P. Samarati. kanonymousdata mining: A survey. In Privacy-Preserving Data Mining: Models and Algorithms, pages 105–136. Springer-Verlag, 2007.

[10] C. Ziemkiewicz and R. Kosara. Embedding Information Visualization Within Visual Representation. Advances in Information and IntelligentSystems, pages 307–326, 2010.

[11] T. Munzner. Visualization Analysis and Design. AK Peters Visualization Series. CRC Press, 2014.