

Visualización de Datos Financieros en la Bolsa de Valores de Colombia

Solano Diana Catalina, Liñán Santiago, Díaz Juan José

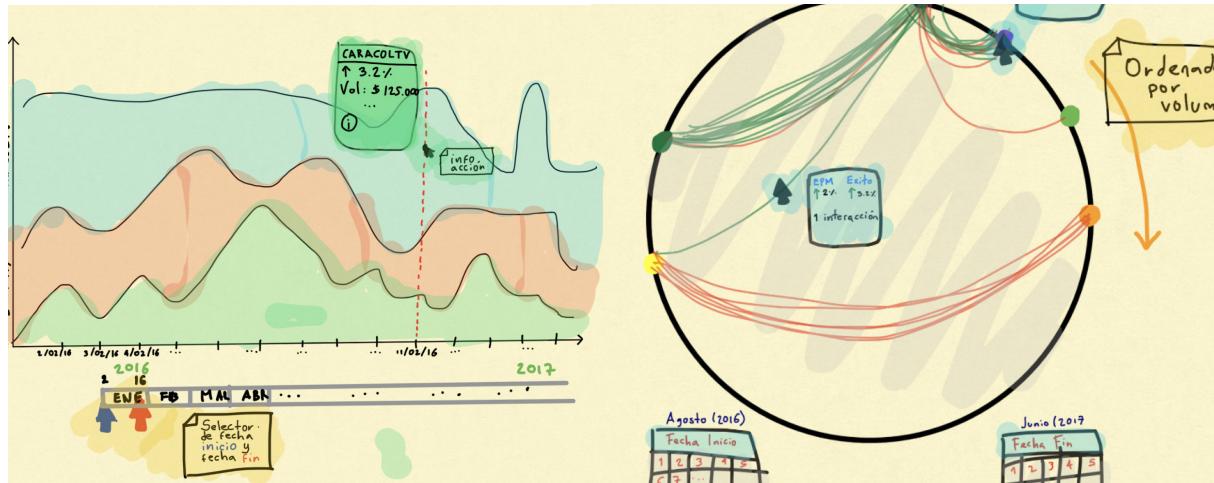


Figura 1. Mockups of some visualization proposals.

Abstract—Among the financial markets a basic goal is to communicate easily and fast the state of the financial stocks that belong to a stock exchange, this is given by the volatile nature of trading and the large amounts of money involved. Nowadays, professional traders and investors have specialized and powerful software to analyze the market, however, beginners in the world of finances go through enormous barriers to understand the market. The purpose of this tool is to support the learning process of new investors in the Colombian stock exchange (BVC), and motivate people with limited knowledge in the finance world to make the first move in the market.

Index Terms—Visualization, finance, stocks, stock exchange.

1. BIBLIOGRAPHIC REVISION

1.1. Importance-driven visualization layouts for large time series data [1]

In this article the authors show a method similar to a treemap to create a time series visualization. This method allows to visualize and compare different time series. In figure 2 there is an example using stock market data.

In this visualization the time series are represented as bar charts in which the height and color indicate the value of the stock in a given time. Besides, stocks with high importance are shown bigger, and stocks of low importance are smaller and located just below the stocks with higher hierarchy.

This visualization is useful to stock market data as it allows to compare many stocks and aggregate many variables, such as transaction volume. It allows us to show a comparative between different industries, thus the user can observe and understand the importance of some industries in the stock market, once the user is intrigued, the visualization is flexible enough to create more analysis.

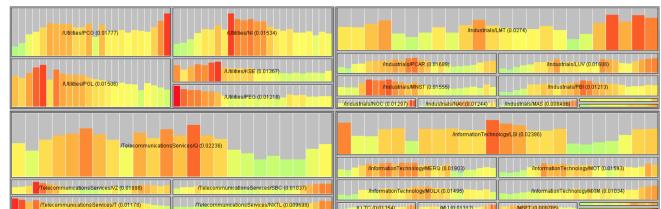


Figura 2. Visualization example of stock market data Hao-Dayal.

1.2. Financial Reminiscences: an example of art based on information visualization in finance [3]

In Saito's article is developed a visualization to identify changes in time of the movements of a set of stocks that belong to a stock portfolio. In order to achieve this, they developed a visualization technique based on textures, using as mark point which represent stocks, and the channel as the color (hue) as the percentage change of the price in a pairs of days, all points inside a square that represents the entire portfolio. We can observe that the most stable portfolios are represented by red, orange and yellow points, which are the lowest percentage changes. On the other hand, the unstable portfolios are compromised by green, blue and purples. In figure 3 are presented the behaviour of a portfolio in two years, it is evident that in the second year the portfolio was more unstable than in the first year. This visualization allows to include time dimension and has high scalability, besides, it facilitates

- Diana Catalina Solano, Universidad de los Andes. E-mail: dc.solano10@uniandes.edu.co.
- Santiago Liñán Romero, Universidad de los Andes. E-mail: s.linan10@uniandes.edu.co.
- Juan José Díaz, Universidad de los Andes. E-mail: jj.diaz1067@uniandes.edu.co.

to incorporate a high number of stocks, which is frequently required with financial data.

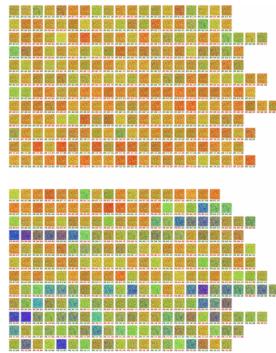


Figura 3. Visualization Financial Reminiscences.

1.3. Casual Information Visualization: Depictions of Data in Everyday Life [2]

In this article is discussed visualization which are targeted to general public, this field is called Casual Infovis, and it has its own study brand, due to its many constraints associated to the user type. According to Pousman et al. one of the biggest difficulties to create a visualization for the general public is to measure how useful it is. The main reason is to how to measure the generated insights a person has obtained using the visualization, and to how measure the usefulness of the information presented in the visualization.

2. TARGET USERS

Using the information granted by Carlos Castillo Arbeláez, Director of electronic media of the Colombia Stock Exchange (BVC), we could identify as target users the people with limited knowledge in stock market, but interest in investment opportunities. These people are usually very curious and motivated, but they are not willing to invest in the stock market due to difficulties understanding finances and stock markets. The objective is to develop an interactive visualization targeted to **people with limited knowledge in stock markets, with the goal of catching their attention and invite them to participate in stock markets.**

3. CURRENT SOLUTION OF THE CLIENT

Currently the client has a visualization in their web platform, this solution is implemented in Flash, which is an outdated technology blocked by many modern web browsers. So an update to a more recent technology such as HTML5 is necessary.

In figure 4 we can observe an example of the visualization the client currently has, this visualization is used for all of their products, from stocks to forex, and consist in a line chart that indicate the variation of the product price as a function of time. It is also possible to select many financial assets to visualize them as other lines in the line chart. The lower part of the visualization shows a brief summary of all the time series of the asset. And finally shows as a horizontal line the base price of the asset currently in the visualization.

According to the interview, there are multiple problems with this visualization, the first being that it is designed for users with vast experience in finance and not for a general public, this is a problem, and is that most experienced users already have a commercial software solution that gives them all sophisticated visualizations they need, and casual users only have the visualization currently available in BVC web page. The second problem is that the line chart does not allow to visualize the stock price behavior in times of economic growth or recession. The third problem is the difficulty to compare many stock assets, e.g., in figure 4 it is not possible to recognize which stocks have most volume of transaction, or recognize periods in which a stock performs better than the others.



Figura 4. Visualization currently available to Colombian Stock Exchange.

4. CHARACTERIZATION OF THE DATA: WHAT

- The given data correspond to the historic log of transactions for each company in the BVC, in the time between January 4th of 2016 until December 29 of 2017
- These data are open and do not contain sensible information
- The data are in table format CSV
- Data contains 35.982 registers and a size of 3.4MB
- Each row represents the consolidated behavior of a stock in a single day

Next is shown the description of the columns in the table:

4.1. FECHA

- **Description:** Date which defines the time of the current row.
- **Data type:** Date - Ordinal, sequential.

4.2. NEMO

- **Description:** Name symbol of the stock in the BVC.
- **Data type:** Text - Categorical.

4.3. RAZON_SOCIAL

- **Description:** Name of the company to which it belongs the stock.
- **Data type:** Text - Categorical.

4.4. VOLUMEN

- **Description:** Volume size of transactions of the stock in the given day.
- **Data type:** Number - Ordinal, quantitative, sequential.

4.5. VAR_PRE

- **Description:** Variation of the stock price in the given.
- **Data type:** Percentage - Ordinal, quantitative, divergent.

4.6. CANTIDAD

- **Description:** Number of stocks sold in the given day.
- **Data type:** Positive integer number - Ordinal, quantitative, sequential.

4.7. PRE_CIE

- **Description:** Close price of the stock.
- **Data type:** Positive number - Ordinal, quantitative, sequential.

4.8. PRE_MAX

- **Description:** Maximum price of the stock.
- **Data type:** Positive number - Ordinal, quantitative, sequential.

4.9. PRE_MED

- **Description:** Mean price of the stock in the given day.
- **Data type:** Positive number - Ordinal, quantitative, sequential.

4.10. PRE_MIN

- **Description:** Minimum price of the stock.
- **Data type:** Positive number - Ordinal, quantitative, sequential.

4.11. V_ACMI_FEC_INSCRIP

- **Description:** Date in which the stock started to trade in stock exchange.
- **Data type:** Date - Ordinal, sequential.

4.12. PATRI_CON

- **Description:** Equity of the company.
- **Data type:** Positive number - Ordered, quantitative, sequential.

5. TASKS DESCRIPTION: WHY

Next is shown the tasks that the visualization to develop should accomplish:

- **Identify outliers** of the stocks with the highest volume of transactions
- **Explore characteristics** of the stock market as a function of time.
- **Compare tendencies** of growth and decline of the stock value, and its volume in the market.
- **Identify correlation** between the movement of price of different stocks in a specified time range.

6. SOLUTION PROPOSAL: HOW

Next, is shown the proposed solution for the tasks. The proposal include a series of the complementary visualizations:

6.1. Treemap

The volume of transaction is calculated with with the selected time period for each stock. Also, using external sources we will obtain the industry sector each company belongs in order to aggregate stock by sector.

6.1.1. Marks

- **Rectangle:** Represent a single stock in a given time.

6.1.2. Channels

- **Color (Hue):** Represents an economic sector.
- **Area:** Represent the volume of transaction of a stock in a given day.

In figure 5 is presented a mockup of this proposal.

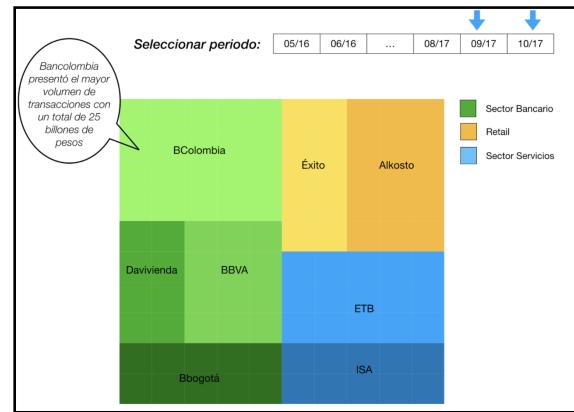


Figura 5. Mockup of the Treemap

6.2. Adjacency matrix

The user must select a time period to visualize. The correlation between all stocks are calculated, on each axis of the matrix are the stocks and the value on the matrix correspond to the correlation between stocks, if positive then its green, if its negative then is color red. The color saturation define the magnitude of the correlation.

6.2.1. Marks

- **Cell:** Represent the correlation between 2 pair of stocks.

6.2.2. Channels

- **Color (Saturation):** Represent the magnitude of the correlation between a pair of stocks
- **Color (Hue):** Red color represents a negative correlation, green represent a positive correlation.

In figure 6 is shown a mockup for this proposal.

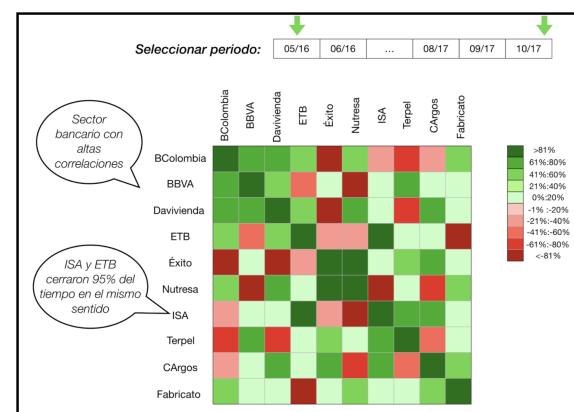


Figura 6. Mockup of adjacency matrix

6.3. Stacked Area Chart

it shows the change in the volume of transactions of different stocks in a given time. The Y axis shows the volume and the X axis is the time axis.

6.3.1. Marks

- **Shape:** Represent a stock.

6.3.2. Channels

- **Color (Hue):** Used to differentiate a stock from others.
- **Height of a shape in the Y axis:** Represents the volume of transactions of a stocks in the respective time in the X axis

In figure 7 is the mockup of the proposed visualization.

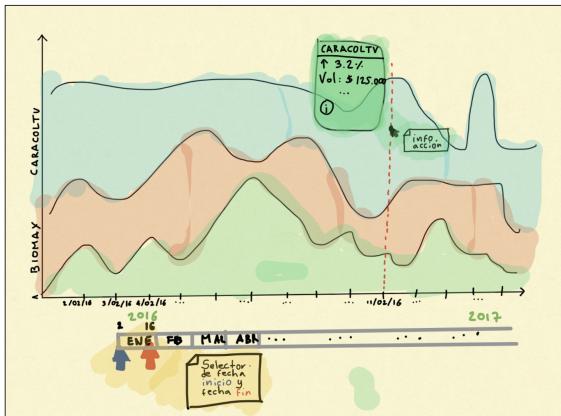


Figura 7. Stacked Area Chart Mockup

6.4. Edge Bundling

Contains in its perimeter the different stocks traded in a specified time. A green link between a pair of nodes represents that both stocks had a positive variation more than 1% in a day. A red link indicates that both stocks had a negative variation in less than -1% in the same day.

6.4.1. Marks

- **Line:** is the interaction between a pair of stocks.
- **Dot:** is a stock.

6.4.2. Channel

- **Color (Hue):** Green and red. Green indicates a positive relation, and red a negative relation.
- **Position in the circle diameter:** Stocks are ordered from highest to lowest according to their transaction volume.

figure 8 shows a mockup of the proposed visualization.

7. MEETING LOG

The meeting took place on october 2nd of 2017 at 4pm in the 12th floor at the BVC (Bolsa de valores de Colombia). All the members of the group assisted to the Carlos Castillo office (Director of electronic media at BVC).

- Firstly, all the group members were introduced. Carlos Castillo introduced himself and shortly described the objective of the project: “communicate easily and fast the state of the financial stocks that belong to the BVC (Colombian stock Exchange)”
- Carlos Castillo showed us the BVC’s web page, in which the data is available. We discussed about how the user interface was not friendly to people without knowledge of finances. We interacted with a Macro Media Flash development which enables the comparison between different stocks all through the day, however, a large section of the data is presented in tables.
- We asked Carlos about examples to present the financial data he gave us. Carlos told us that a great portion of platforms (like Bloomberg), are oriented to experts in the topic.

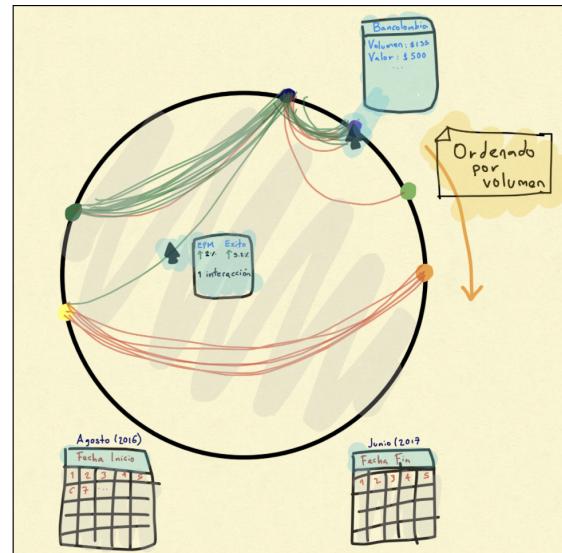


Figura 8. Edge Bundling Mockup

- Carlos gave us the historical transaction log of stocks between 2016 and 2017. Data was cleaned and filtered.
- We were instructed about the cells, columns and rows of the database. We asked some questions about the bourse main activities. The meeting ended.

The meeting ended at 17:10. Carlos sent the database to the members of the group.

8. USER VALIDATION

3 interviews were made to potential users. Final users where people without knowledge in finances, but interested in trading eventually in the stock exchange. People with different ages and professions were chosen.

8.1. Interviews

8.1.1. Interview #1

Woman, 52 years, Psychologist

- The diagrams had to be explained in detail. She did not understand the adjacency matrix.
- In the tree map is not possible to understand which stocks are being evaluated. She would like to have more information about the stocks.
- In the adjacency matrix, she would like to know the financial sector in which each stock is involved.

8.1.2. Interview #2

Man, 23 years, Master student in Industrial Engineering

- He would like that the Edge Bundling graph could present the negative and positive correlation. He would like to filter the information by financial sector and by name.
- In the treemap, he would like to filter the stocks by volume.
- Filter more information in the adjacency matrix (financial sector, volume, value of the stock, etc).
- He would like to hide stocks in the Stacked Area chart.

8.1.3. Interview # 3

Man, 40 years, Physician

- Did not understand the Edge Bundling mainly because theoretical concepts (correlation between actions).
- In the treemap is hard to compare similar areas, Areas should be ordered.
- The colors of the adjacency matrix are too strong.
- In the stacked graph it is too hard to compare stocks.

8.2. Conclusions about the interviews

- Diagrams should be more descriptive and clear.
- It is necessary to explain the finance concepts to the users, given the fact that they are not experts in the field. (Volume of transactions, correlation, risk, return).
- it is useful to include interactive examples.
- Different filters must be implemented to facilitate the comparison between stocks. Should include volume, sector, etc.
- Include tops (the top 10 more traded, the top 10 riskier)

8.3. Usability tests developed during the class

8.3.1. Edge bundling

- It should show negative correlation. Furthermore it must indicate the value of the correlation presented in the graph.
- Allow the search of a specific stock.
- Include a help menu, which prevents that the user have to chase the tooltip with his eyes.
- Improve the usage of colors, a green correlation color not necessarily represents something good. Also, search friendlier colors with colorblind people.

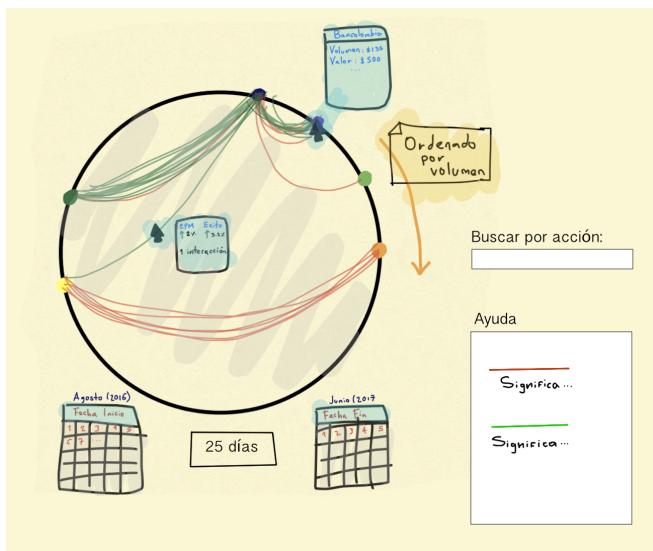


Figura 9. Corrections made in the Edge Bundling

8.3.2. Treemap

- The used color can't be interpreted, one possible alternative is to indicate the cost effectiveness of the stock over the time.
- It is hard to differentiate the size of the volume of transaction when there are 2 sectors of similar size.
- It should be shown the behaviour of a stock when an user wants to know details of a stock.
- To improve the date selections, date pickers and calendars should be used.

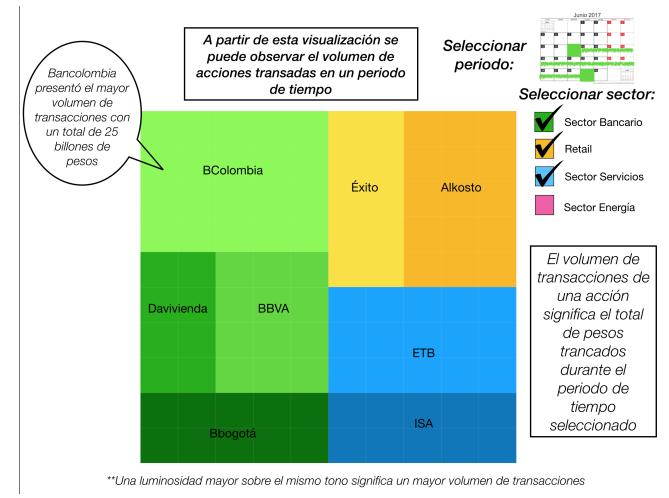


Figura 10. Corrected Treemap

8.3.3. Adjacency matrix

- Explain the concept of risk diversification
- The graph is very intuitive

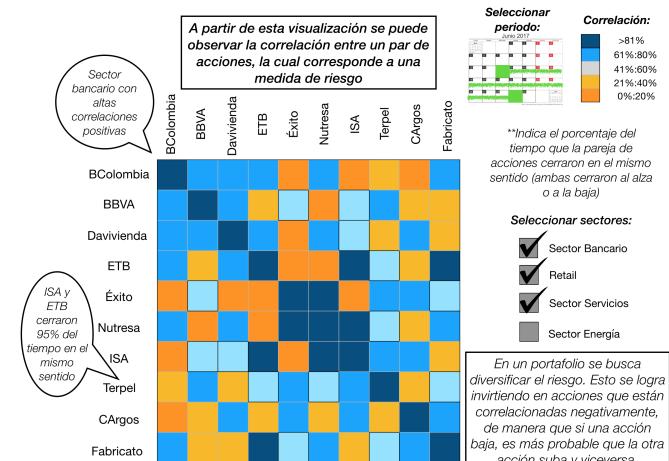


Figura 11. Second version of the adjacency matrix

8.3.4. Streamgraph

- Show the results in clear tooltips. Focus on identifying days and moments of high volume of transactions.

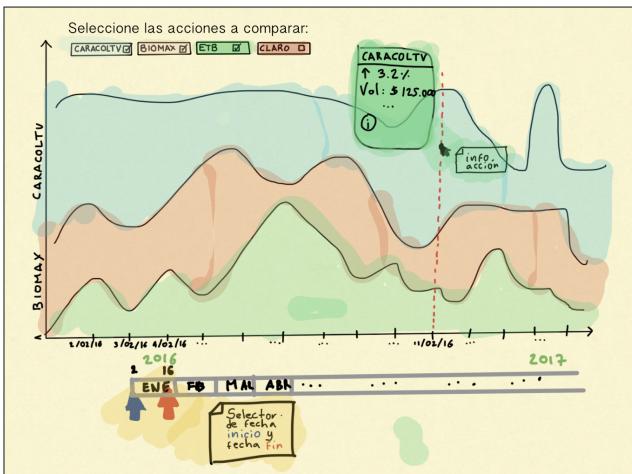


Figura 12. Corrected Streamgraph

8.3.5. Grading pf the users during the usability tests

Question: from 1 to 7, being 7 very good and 1 very bad, how would you grade these prototypes?

Interview 1	5
Interview 2	6
Interview 3	6
Average	5.66

9. RESULTS

The visualizations were developed entirely in D3 ??, a popular Javascript framework for visualizing data in HTML, using SVG and CSS. The final product can be found at <https://slinan.github.io/BVC>.

The visualizations will be presented, including the relevant considerations for each.

9.1. Treemap

The treemap was developed, taking into account the advice given by the final users. Different colors represent the different financial sectors of the economy. There are date selectors which help to modify the time lapse. The user can click on an economy sector; after that, he will see the biggest stocks traded in those dates. Additionally, a line chart of the stack will be shown to see the behaviour during the analysis time. The developed treemap is presented in the figure 15

Treemap represented a great success, and users were able to identify the stock transactions with more volume in the industry. Furthermore, when the user hover on a specific stock, a historical line chart of the stock is shown. This allows the user to identify the moments in time that leaded to the area shown whit in the graph.

9.2. Stacked area chart

The stacked area chart allows the user to analyze and determine specific moments in time, where stocks went up or went down. We graph the volume of transactions during a selected period. This visualization is useful to identify the periods where the volume of transactions of a selected stock increases or decreases. Also, an user can compare multiple stocks over the same period and, understand its behavior.

9.3. Adjacency matrix

The adjacency matrix its a useful visualization to identify the relations between any pair of data. We graph the stocks with transactions during a year and we calculated a metric of relation between stock. The metric was the number of times that the pair of stocks closed both increasing or both decreasing. We included this metric as a measurement of risk. We found out that there aren't relations between the stocks, this is a consequence of the Random Walk Theory, which states that the value of the stocks follows a random distribution.

Volume of transactions by stock sector

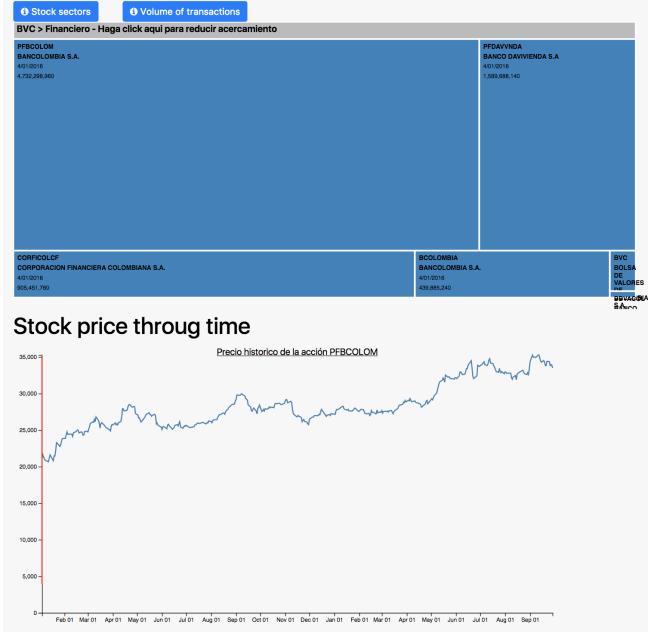


Figura 13. Developed treemap

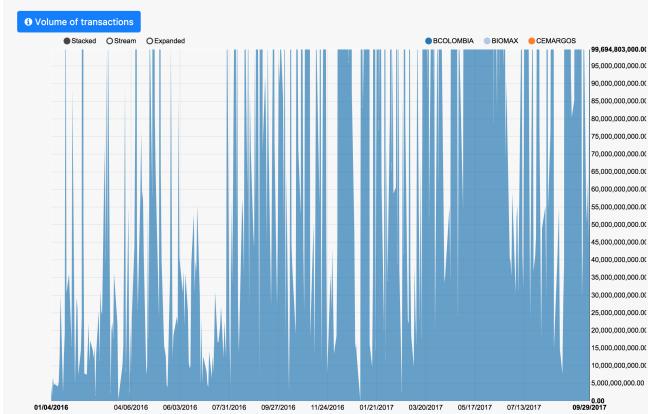


Figura 14. Developed Stacked Area Chart

9.4. Edge bundling

We proposed an edge bundling visualization as part of the project proposal. Even though, from the adjacency matrix we found out that there is no correlation between the stocks in the BVC, that is why we decided to omit this visualization. Instead of the development of this, we developed a detailed tutorial including the explanation of each visualization and the definition of some important finance concepts.

10. INSIGHTS

Some insights that a user with limited knowledge in finance could find in our application would be:

- Dependencies between stocks are completely random and can not be predicted (Random Walk Theory).
- The sector which the highest volume of transactions is finance.
- There is a large number of stocks in Colombia that does not change even in a couple of months.
- Big companies like ARGOS or BANCOLOMBIA move thousands of millions of pesos in a single day.

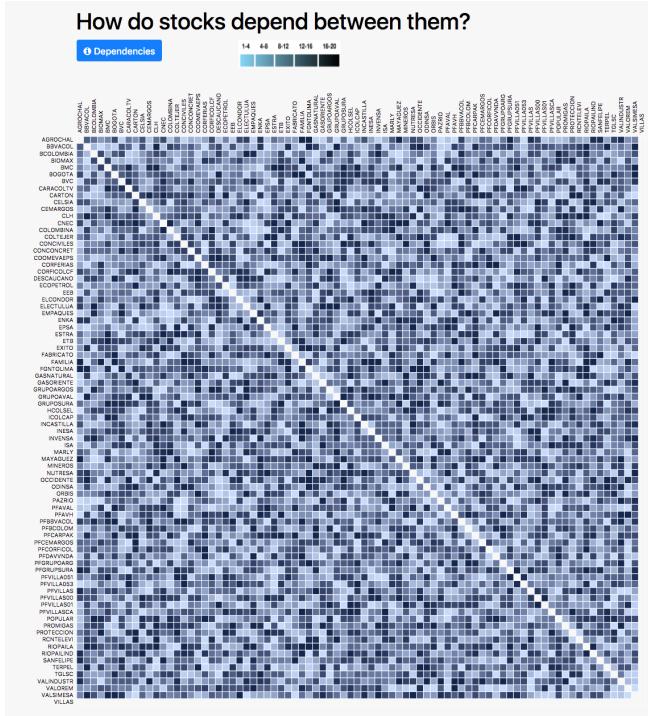


Figura 15. Developed Adjacency Matrix

11. CONCLUSIONS

- The process of designing a visualization must start from problems and users. From this point, the Tamara's framework guides the understanding of the problem and the development off a solution.
- The usability tests were useful to understand how people interact with the visualizations; and determine how to improve them to successfully accomplish the defined tasks.
- The way the financial market presents stocks information, is not designed for people with limited knowledge in the field. Usually graphs are difficult to understand, and there are not definitions of the financial concepts that are involved.
- The development of the project included a tutorial, which transforms paradigms that defines how financial information is presented.

REFERENCIAS

- M. Hao, U. Dayal, D. Keim, and T. Schreck. Importance-driven visualization layouts for large time series data. *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005*.
- Z. Poussman, J. Stasko, and M. Mateas. Casual information visualization: Depictions of data in everyday life. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1145–1152, 2007.
- Y. Saito. Financial reminiscences: an example of art based on information visualization in finance. *Proceedings Sixth International Conference on Information Visualisation*.