

Visual Comparison of groups explaining a time event

Juan Sebastian Cardona*
Universidad de Los Andes

John Alexis Guerra-Gomez†
Universidad de Los Andes
UC Berkeley

ABSTRACT

The objective of this project was offer the users a tool that allows them to generate time events and find relationships between variables. The project was divided in two: -the beta case which was made specifically to find the re-hospitalization time event relations with the other variables. - The second part was to make a general application with a UI that allows the user to generate their time events. We work hand by hand with the investigation specialist of the Hospital Militar Adriana Beltran, who helped us giving us feedback of the application, such as "make the category comparison optional, because there are users that would only want to analyze one category". Finally, there will be future work, such as improve the UI, and the way in which the bins of the temporal and quantitative data. You may find the implemented tool and check the code at <https://dcptool1.firebaseio.com> and <https://github.com/jscardona12/hospitalviztool> respectively.

Keywords: Data Analysis; Data Comparison; Data Processing; Visual Analytics; Visual Comparison; Time Events

1 INTRODUCTION

Visual Analytics is an important topic that all companies should take into account since nowadays those companies have a big amount of data. However, in most cases they don't know what to do with all that information. Given the above, nowadays there are a lot of companies who are dedicated to visual analytics. Most of the solutions we have today are oriented to visualize all the dataset. Nonetheless, they don't allow the user to generate time events and make a visual comparison to understand it or to compare different parts of the data in order to identify relations between variables. For this reason, the purpose of this project is to develop a web application that allows a user to generate a time event, compare their data and understand variable relations, by contrasting different parts of the dataset, in a intuitive way.

The rest of this paper is organized as follows: we introduce the general and specific objectives of the tool to build. An analysis of the state of the art in the current visualization tools and the ways on how the hospitals are using their data. A proposed solution is made to show the stages and requirements of this project. Next we present the implementation of our approach, this contains the architecture and the technologies used. After that, we describe the sample dataset given by the Hospital Militar Central. That description is followed the test made to validate the solution and the results we obtain. Finally we discuss the current work and present the work to be made.

2 OBJECTIVES

2.1 General Objective

Allow users to create and understand a time event in an intuitive way, finding relations between variables by data comparison.

*e-mail: js.cardona12@uniandes.edu.co

†e-mail: ja.guerrag@uniandes.edu.co

2.2 Specific Objectives

- Develop a script which allows to know if a patient has been re-hospitalized
- Develop a Beta version of a Web Application that allows the Hospital Militar Central to load data and visualize the data .
- Perform beta testing.
- Develop the final version of the Web Application, that allows the Military Hospital to understand the data, through a variable comparison functionality.
- Develop a general Web Application that anyone can use to understand relations between variables.
- Launch the application as a free SaaS.

3 STATE OF THE ART

Nowadays most visualization tools are made for developers, this means that a person that doesn't know anything about data will find it very difficult to use.

On the other hand, the use of hospital data are more focus on event prediction with machine learning, than on understand the data. Having into account that event prediction is very use full, in machine learning you can't know which are the variables that make that event happens, because you don't know how the model is being trained.

- Google Analytics[4]: is an Analytic tool that offer free and enterprise analytics tools to measure website, app, digital and offline data to gain customer insights.[4]

The pros of this tool are that it is an automatic tool, you only have to paste a script in order to have a visualization of your web page. It have a lot of functionalities that allow the user to understand what is going on with the web page.

The cons of this tool are that it only works with webpages and you cannot upload your own data to visualize it.

- Voyager 2 [3]: is a data exploration tool that blends manual and automated chart specification. Voyager 2 combines PoleStar, a traditional chart specification tool inspired by Tableau and Polaris (research project that led to the birth of Tableau), with two partial chart specification interfaces: (1) wildcards let users specify multiple charts in parallel, (2) related views suggest visualizations relevant to the currently specified chart. With Voyager 2, we aim to help analysts engage in both breadth-oriented exploration and depth-oriented question answering.[3] The pros of this tool are that it gives you the best visualization depending your type of data, you can upload your data to be visualized and it is open-source.

the cons of this tool are that it doesn't give the user what it needs to understand variable relationships. It is oriented to developers, not oriented to common users.

- Tableau[6]: is a Data Analysis to where you can upload any type of data, and allows you to create graphs, tables, maps etc. in order to help you understand the data.

The pros of this tool are that you can upload your data to be visualized, it has a lot of options to visualize your data.

The cons of this tool are that it can be difficult to a person that is not a developer, because you have to know what kind of data you have and in what ways you can use it. It is also not free, so you have to pay a license.

- Graphext [5]: is an advanced analytics solution to help businesses make better decisions based on data. We capture their data, enrich it and provide a powerful visual interface to find actionable insights about their customers, market and products[5].

The pros of this tool are that it allows the users to understand the data and use that information, you can also give your data to them so they can analyze them.

the cons of this tool are that it is a demo version, Graphext makes a restructuration of your data.

Although these are well known intuitive and used tools, neither one has a comparison module that allows select parts of the data in order to have a better understanding of the variable relations, also they don't have a module to create time events.

On the other hand, there are solutions [7][8] oriented to help hospitals by using machine learning to predict re-hospitalizations and possible deaths using the hospitals data, or even to detect diseases faster.

4 PROPOSED SOLUTION

As previously stated, the goal is to create an application that allows any user to easily identify and understand relations between variables. The application will be done in 2 phases.

The first one is build a specific tool for the Hospital Militar Central. In order to build this tool, is important to develop a script that detects the time event "re-hospitalizations" and a web application that will allow the hospital to know how many re-hospitalizations they are having and which variables are the most sensitive about this. The Application will allow the user to select different parts of data to compare it. Also, they will be able to analyze all the other variables, such as hospital fees.

The second one is an open source application. This application will allow every user to charge a dataset and create a time event getting an intuitive visualization that allows them to make a comparison of all the variables and the categories of the variables, allowing the user to find relationships between the variables and the time event or the other attributes.

4.1 Solution Requirements

- Functional Requirements:
 - Allows CSV file upload.
 - Creation of time events: A time event is defined as: an event that happens between a start date and an end date , within a period of time(days) for a same identifier.
 - Main variable selection.
 - Distribution Graph of the main variable.
 - Main category selection.
 - Show the sensitivity of the variable to the main category.
 - Secondary variable selection
 - Show relationship graph between the main variable category and the secondary category.
 - Comparison of categories of the main variable.
 - Comparison Graph of categories of the main variable.

- Non Functional Requirements:

- Intuitive User Interface.
- Assure that the user can make every action in less than 8 clicks.
- Calculate the time event in less than 5 minutes.
- Availability 99.95% of the time.
- Deployed Web Page.

5 DCP TOOL

DCP allow you to understand the behavior of a variable, find relations between variables and make comparisons to understand the properties of an specific relation.

5.1 Architecture

To make things easy to the user, we decided to implement all the application server-less, this means that the application is entirely on the client. The Application is deployed with Firebase hosting.

5.2 Technologies

For the web implementation we use the Facebook framework React, HTML5 and CSS3. For the graphs and data processing we use the library Vega , specifically Vega-Lite, this library is built over D3js. For the UI design we use the library Material-React and ANTD for the file dragger and loading gifs.

6 DATASET DESCRIPTION

The recollected data was given by the Hospital Militar Central, they gave us a file with 65000 registers that contain the data of the hospitalization process. After they give us the data, we transform the data to a CSV file . With this data we create the time event where the starting date was the field "FECHA_INGRESO" the end date was the field "FECHA_EGRESO" the identifier was the field "id" and the period was thirty days.

7 TESTS AND RESULTS

For the validation of the solution, we worked hand by hand with the domain expert Adriana Beltran and with the directors of the Hospital Santafe.

7.1 Tests

the researcher of the Hospital Militar Central, she gave feedback such as "Put the comparison as optional, sometimes the users don't want to compare", she gave us positive feedback saying that she never imagined that the people who are most re-hospitalized are those in their twenties. the directors of the Hospital Santa Fe said that the application was very good in terms of usability and it allows to find the relation between variables very easily.

Thanks to this feedback, we improve DCP making more intuitive the comparison module and doing it optional, we also implement interactive graphs to show all the data in only one graph.

7.2 Steps to use the application

Below are the steps of how to use the application.

1

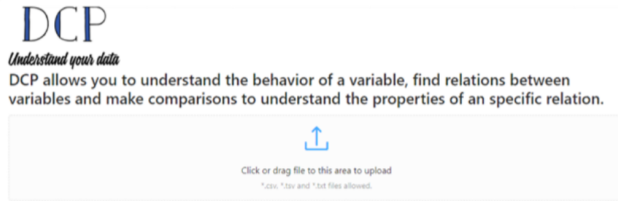


Figure 1: Main Interface: In this interface you can upload your CSV file.

2

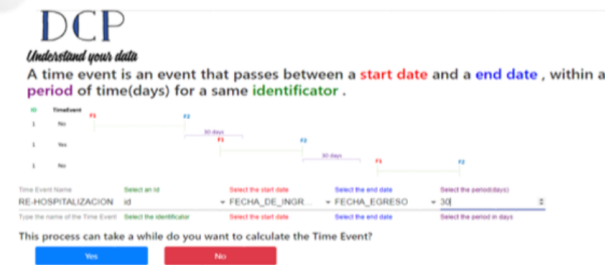


Figure 2: Time Event Generation: In this interface you can create a time event based on the variables you choose.

3

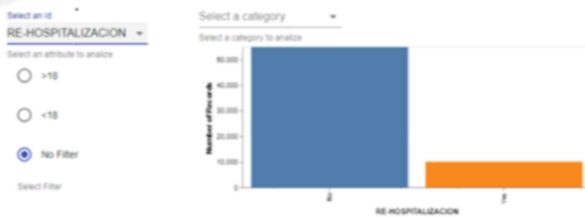


Figure 3: Select a variable to analyze: In here you can select a variable to analyze the behavior of their values and also filter them.

4

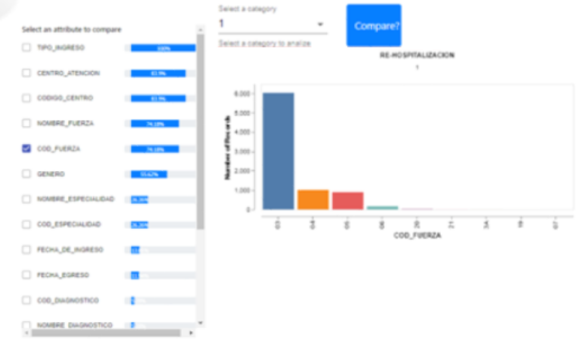


Figure 4: Showing relations between variables: Once you select a category of the variable to analyze, on the left side will appear a side menu showing the percentages of relation between the chosen category and the variable

5

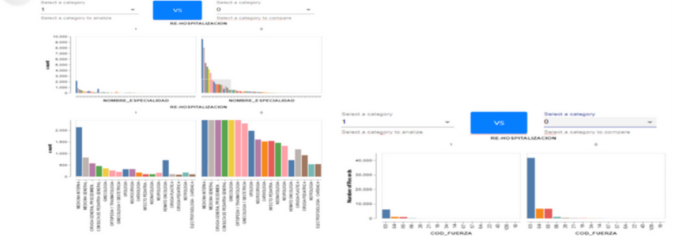


Figure 5: Comparing categories: Once you've selected a variable to know the relation, you can compare the chose category and other category. This interface let the user understand all kind of data, if there is many data, two graphs will appear, the one above to select the interval that will show the one below.

8 CONCLUSIONS AND FUTURE WORK

Usability tests with expert have shown that the application works as expected, nevertheless we need to improve things, such as:

- Make more efficient the bin calculation of quantitative and temporal data.
- Implement filters for all the variables of a dataset.
- Improve the UI.
- Find a more effective way to measure the relevance of a category in a variable.
- Implement users, and a module to store and choose previous converted data.

In conclusion, to solve the difficulty of understanding time events and finding relations between variables, we implement DCP tool, which is a tool where you can upload your own data, create a time event and based on comparisons you can find relations between variables and understand time events. Also, based on the feedback gave by the Hospital Militar Central and the Hospital Santafe, the project was a success, however there is a lot of work to do to implement the new functionalities such as user management.

REFERENCES

- [1] Wongsuphasawat, Kanit Moritz, Dominik Anand, Anushka Mackinlay, Jock Howe, Bill Heer, Jeffrey. Voyager: Exploratory Analysis via Faceted Browsing of Visualization Recommendations.

- [2] Taylor-Powell, Ellen Renner, Marcus, C. (2003). Analyzing Qualitative Data . Retrieved from pubmedcentral: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2474845/?tool=pmcentrez>
- [3] V. (2018, February 18). Vega/voyager. Retrieved February 24, 2018, from <https://github.com/vega/voyager>
- [4] Google analytics Solutions - Marketing Analytics & Measurement. (n.d.). Retrieved February 24, 2018, from <https://www.google.com/analytics/>
- [5] G. (n.d.). Customer Success Analyst Application Form. Retrieved February 24, 2018, from <https://graphext.typeform.com/to/CZVOCW>
- [6] T. (n.d.). Tableau Software. Retrieved February 24, 2018, from <https://www.tableau.com/>
- [7] Alvin Rajkomar, Eyal Oren, Kai Chen, Andrew M. Dai, Nissan Hajaj, Michaela Hardt, C. (2018). Scalable and accurate deep learning with electronic health records . Retrieved from Digital Medicine
- [8] Callahan, Alison Shah, Nigam H., C. (2017). Machine Learning in Healthcare . Retrieved from Key Advances in Clinical Informatics: Transforming Health Care through Health Information Technology