Visual Analytics Tool for Incident Management in a Public Transportation System

Laura Andrea Cortés Fernández, Anamaría Irmgard Mojica Hanke and Meili Vanegas Hernández

***Abstract*—A visual analytics tool is presented which aids a public transportation operator in creating insights while manipulating and interacting with their incidents management data. Based on a time filter, it is possible to discover which day in the past few weeks had an exceptional average time in ticket solving and which are the principal causes that explain that behaviour. Multiple views are presented, allowing interactive data analysis. The proposal is implemented as a NODEJS application, which interacts with a POSTGRESQL and MYSQL databases. The final product aims to fulfill high usability standards and allow intuitive user interaction.**

Keywords: visualization, analytics, incidents.

1. Introduction

This work is presented as the final project of the Visual Analytics course in Universidad de los Andes [1]. Our proposal aims to solve a real problem through a visual analytics application. Along this project we will be working with Recaudo Bogotá S.A.S, a public transportation operator operating in Bogota, Colombia.

Operational problems, also called decision theory, are highly studied, because their high complexity (np complete) and the purpose of them: finding an optimal solution to a problem. Recaudo Bogotá has an operational problem, in the moment of solving tickets (incidents) in an optimal time. Because of the complexity, already mention, the company of transportation and freights needs more computational resources and time, to find an approximate solution. Therefore, we propose a visualization approach, in which an expert can manipulate and create insights with the incidents.

2. Related Work

This section briefly reviews other works about operational and transportation problems. Although there are many ways of approaching, we distinguish two main rapprochement, traditional computation, which need a lot of resources and with the use of visualizations, which needs an expert of the topic that is displayed.

A. Visualization

1) Incident Management Systems: .
There are several studies around the optimal representation for real-time incident management. In public transportation scenarios, the most common research topic is traffic incidents, such as heavy traffic, accidents and breakdowns. This area of study need to manage different attributes for each record, as well as its associated spatial information.

According to the description above, it is important to mention the proposal made by Anwar et al. in which each incident is classified by his type and represented differently [2]. In this case, the spatial data is critically important, because its intentions are primarily to represent its consequence in a determined zone, as it is shown in Figure 1.

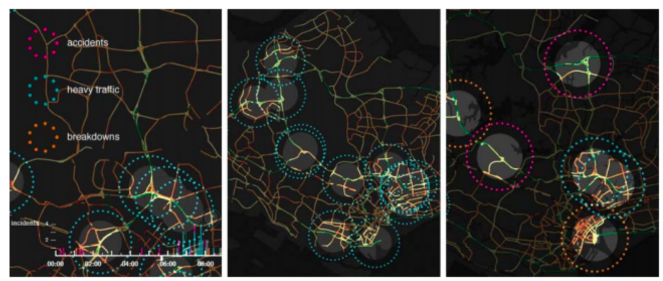

Fig. 1. Traffic incidents highlighted using the Traffic Origins approach. When an incident occurs, its location is marked by an expanding circle that reveals traffic conditions in the immediate vicinity. (Taken from Traffic Origins: A Simple Visualization Technique to Support Traffic Incident Analysis)

Another example was proposed by VanDaniker in 2009 [3]. In behalf of the research topic of interest, he made an investigation in which he proposed relevant methodologies used to represent real-time traffic incident data. Also, he presented the Transportation Incident Management Explorer (TIME), an useful tool that provides support visualizing real-time and historic traffic incident data.

Lastly, Luz and Masoodian proposed temporal mosaics as an alternative to traditional timelines to represent temporal data [4]. However, it has important restrictions when instantaneus events are handled.

2) The pressure cooker of the public transportation in Bogotá: .
The purpose of this project, made by Ardila Gomez in 2006 [5], was to present the oversupply of vehicles in the public transportation system, which causes many complications for the inhabitants of Bogotá in their daily life, and an possible approach justify by visualizations.

To expose better the problematic Ardila presents four main aspects: first an inflated rate; third arrangement made under the table between with the government to assign the distribution of the vehicles for public transport; fourth brings new staff for the same roles that already exists. Based on the above, he presents a solution founded on the strengthen of the government also with a centralized system to collect money and change the freights system.

As previously mentioned, to justify the solution were used visualizations. In this specific case was used a graph to visualize some statistics about the average of passengers in minibus and micro bus in by day. (See Figure 2)

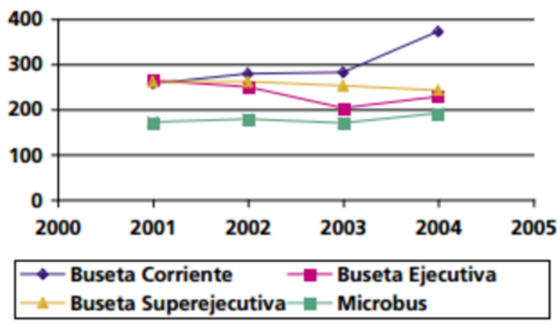

Fig. 2. Passengers by micro bus and minibus. Source: DANE data and calculations of the author (Taken from: Ardila, La olla a presión del transporte público en Bogotá)

This figure, represents data collected by the author and some calculations made by him. In the next figure (See figure 3) it is possible to visualize some data corresponding to big buses.

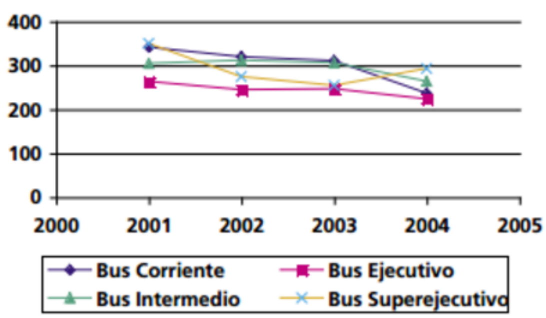

Fig. 3. Passengers bus. Source: DANE data and calculations of the author (Taken from: Ardila, La olla a presión del transporte público en Bogotá)

In the visualizations presented in Figures 2 and 3, it is possible to have an idea of the oversupply of the buses, also how the demand of public transportation has dropped over the years. Therefore, the visualization lacks precision in Y-axis, which does not allow to the user, expert to know the truly data. Moreover, one channel that was used (*color hue*), was not the most appropriate, because some colors are difficult to identify. Another thing to improve is the scale, because it does not allow the user to view the real impact of the dropped of the demand over the years.

3) LIVE Singapore!: .
The purpose of LIVE Singapore! [6] was to use data in real time for creating tools that helped citizens to make choices during their daily activities. The center of the project was to interpret big data through visual representation.

Consequently, it is used spatial graphs representing data, such as, cab's journeys to show the time that takes to move through the city. Then, to represent it they use the size of the map to represent traffic, each area change size independently, if the time that takes to move increase, the map gets smaller.

Therefore, the channels used to represent attributes (movement and area) are not the most suitable. Firstly, the movement does not allow to identify how much the traffic decreased with precision. Furthermore, the fact that each area is independent, it loose highlight from the visualization.

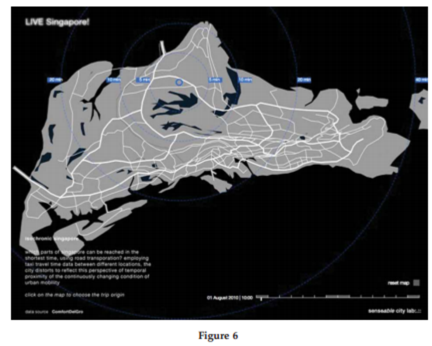

Fig. 4. Source:Journal of Urban Technology, April 2012

B. Traditional computation

1) Emergency, contingency and incident management system and method: .

The purpose of this patent[7] is to implement a software system and methods to implements real-time management of events considered as incidents,contingencies and emergencies by responding to environmental detectors and user inputs providing constant monitoring, recording data and communication facilities between personnel and coordinators.
Consequently for this type of monitoring and recording data a lot of computational resources is needed.This is because the methods that implements regular monitoring to the user inputs, and the processing of this information for establishing the corresponding targeted communication channel. An example of this usage of large amount of computational resources is the network system that is used, that has local and remote databases, servers and computers. (See figure 5)

As was mentioned the traditional computation use a big amount of resources. It is because the task to monitor and analyse data to find an optimal solution for establishing a communication channel needs high computational power.

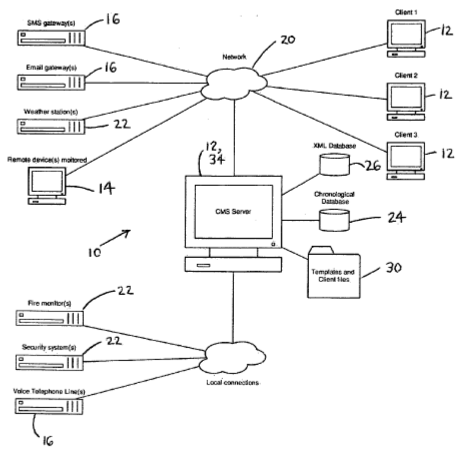

Fig. 5. Source:Emergency, contingency and incident management system and method,2005

3. Description of the Visualization

In the next figure 6 it is possible to see the process followed, in broad outline, by us to construct the visualization. In all the process we try as much as possible to have contact with Recaudo Bogotá.

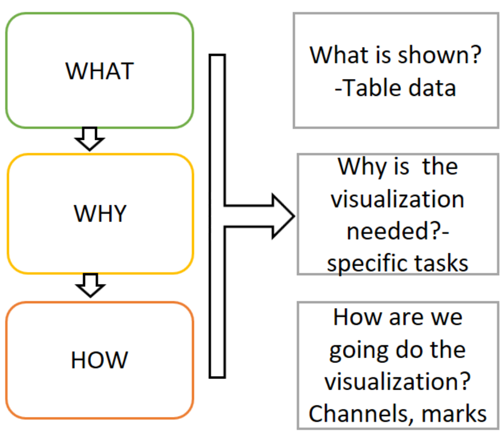

Fig. 6. Process followed to find tasks and how the visualization was going to be made

In the first part of the process (WHAT) after understanding the problem that Recaudo had solving incidents, also called *tickets*, we had to identify the type of data that we would be working on. At the end of the process we identify that was a data that could be represented in a table. Also we had to analyse the quality and the amount of data that we wold be working with.

In the second step (WHY), we had to identify some task that would solve the problem or at least relieved as much as the visualization allows. In this process we identify six different tasks:

T.1 **Present trends** of the total response time of all tickets (by type of incident).

T.2 **Compare similarity** of the average response time, for each state in a defined range of time (week, day or hour), with the expected, which is the exact same range in the past month

T.3 **Identify outliers** from the response time in a defined range of time.

T.4 **Compare distribution** of time for different tickets choosing the combination of states, based in the time that is selected in the filter.

T.5 **Look-up** a single ticket with the corresponding delay between each of the different states.

In the last step (HOW), the principal aim was to find the best visualization, marks and channels, to illustrate the data obtained and accomplish as much as possible tasks. In this process it was necessary to iterate more than one time to find a good approach to tasks.

In this process it was needed to be aware of the information that was analyzed in the second step.

- Each ticket has a unique number.
- The type of data (dates, timestamps, numbers).
- Recaudo, which has problems with "management incidents / tickets" has a process in which a ticket goes trough different states. This process is not linear, it means that before a ticket is created it can go trough the same state more than once (See figure 7).

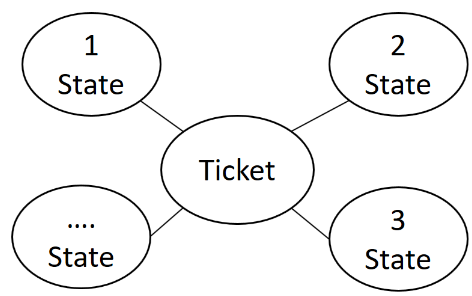

Fig. 7. Process followed to find tasks and how the visualization was going to be made

Based on these key points, we were able to identify that a chart that does not allows a cyclic process between states was not a good choice. Also, the type of data restricts the graphics that can be done. For example a pie-chart is useless with the ticket id and the different states. Another important thing is to be able to differentiate the states along the visualization, also to compare between dates or be able to see a time-line with the distribution of the tickets.

Taking into account all of that was mentioned we decided that the visualization, was going to have different views.Each view was going to accomplish a different task and to complements between each other.

In the first view we implemented a time-line than can be zoomed. In this time line is explicitly mentioned the difference in time between now and the date in which was issued. This view acts as a filter for the other views.To explained it better the user can choose some days, and the tickets of these days are displayed in the other views (this view is located at the

top of the visualization).
In the second view (located under the first, at the right), is a stacked-bar chart that allows to see in the data selected, the corresponding time for a state for each ticket that is displayed.
The third view (located under the first, at the left), is a scatter plot combined with three discontinuous lines. The purpose of these lines is to show the maximum, minimum and average value for the time in one state, meanwhile the scatter plot show the states (of each ticket) by time in each of one.
The next view is a table, in which the user can choose a maximum of five different tickets from the scatter plot and all the specific information (for these tickets) is displayed.

4. Experiments and Results

The resulting visualization is presented in the Figure 8. This tool consists on four specific graphs. Each graph has a different application and pretends to fulfill the user's task, which is, identify which was the day of the previous week in which they lasted long solving incidents. The first visualization is a filter, with which the user can select the relevant days he desire. This filter allows to identify the behaviour of incident management in the past days. Each day is represented with a point and the position of the point represents the average duration of the incidents solved during that specific day. In addition, the day of the week for each day is represented with different color hue, in order to classify each day in week day or in weekend day.

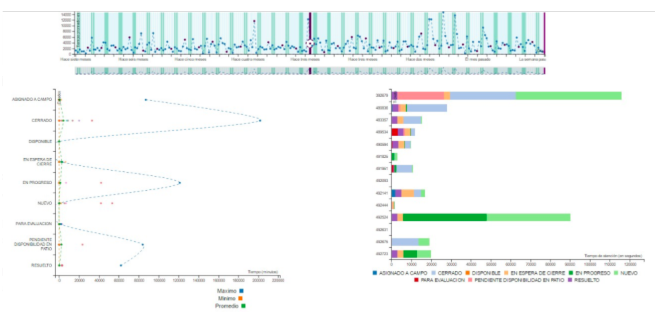

Fig. 8. Principal visualizations in resultant tool

The principal functionality of the filter is to select the day or days of interest. When a day is selected, the Visualization 2 and 3 are updated. Additional, when a day is selected, the corresponding points in the filter graph that have the same week day value as the selected change of color. This functionality allows the user to see an overview of the behaviour of that specific week day over time.

5. Conclusions and Future Work

Visualization with big amounts of data, variables and attributes of different kinds always introduce a high level of complexity. This same behaviour was identified during this project, taking into account the different ranges and domains in data.

The resultant approach presented in this project has strengths and flaws, but it was design in order to maximize the advantages, considering the problem's context and the final user's intention.

It is important to emphasize future work in the following three different aspects: visualization quality, data integration and user validation. In the first case, it is important to present as a result a tool that maximizes the effectiveness. However, during the implementation, there were several limitations with the API's management that restrict the possibilities to maximize the effectiveness.

In the second place, there were multiple difficulties with the data integration. Possibly, there were much more functionalities to explore, taking into account the *dataset*, but at the data preprocessing stage, multiple facts delay the data examination.

At last, there were very few validation stages, in which the tool was examined along the user. This situation affected directly to the project's results. It would be highly recommended to increase the user-validation sessions in which the visualization is examined and evaluated.

References

[1] U. de los Andes, "Isis 4822 : Visual analytics." http://johnguerra.co/classes/isis_4822_fall_2016/, 2016.

[2] A. Anwar, T. Nagel, and C. Ratti, "Traffic origins: a simple visualization technique to support traffic incident analysis," in *2014 IEEE Pacific Visualization Symposium*, pp. 316–319, IEEE, 2014.

[3] M. VanDaniker, "Visualizing real-time and archived traffic incident data," in *Information Reuse & Integration, 2009. IRI'09. IEEE International Conference on*, pp. 206–211, IEEE, 2009.

[4] S. Luz and M. Masoodian, "Visualisation of parallel data streams with temporal mosaics," in *Information Visualization, 2007. IV'07. 11th International Conference*, pp. 197–202, IEEE, 2007.

[5] A. A. Gómez, "La olla a presión del transporte público en bogotá," *Revista de ingeniería Universidad de los Andes*, no. 21, pp. 54–65, 2005.

[6] K. Kloeckl, O. Senn, and C. Rat, "Journal of urban technology fecha: April 1, 2012," *Urban Technology*, vol. 19, no. 2, pp. 89–112, 2012.

[7] J. Douglass, D. Wallis, N. Martyn, and A. Pearse, "Emergency, contingency and incident management system and method," 2005. US Patent App. 11/082,189.