SPLOR-T

An Interactive Urban Spatiotemporal Data Exploration Web Approach

Author:

Miguel Alfonso Feijóo García

Research work to obtain the title of:

Master's in information engineering

Advisor:

Prof. José Tiberio Hernández Peñaloza, PhD

Universidad de los Andes

Bogotá D.C., Colombia

UNIVERSIDAD DE LOS ANDES

FACULTY OF ENGINEERING

DEPARTMENT OF SYSTEMS AND COMPUTING ENGINEERING

MASTER'S IN INFORMATION ENGINEERING

BOGOTÁ D.C.

May 2019

Table of Content

Table of Figures

Table of Tables

Acknowledgments

This research thesis culminates thanks to the contribution of several people who gave me their support during its realization.

First of all, I thank God, because it gave me the opportunity to study and complete the Master's program in Information Engineering at the Universidad de los Andes. To him, because I was always accompanied by him helping me and guiding me in all the processes carried out to reach this precious and willed moment. In a very special way I thank my family, who is the pillar of my formation. I thank my father José Genaro and my mother Alix with their own name, for all their unconditional support, for their love and constant advice. I also thank in a very special way my older brother Pedro Guillermo for all his support, advice and motivation in my development not only of this research Thesis, but professional. Finally, to my grandmother Maria Alix, who I have the certainty that you helped me from heaven and gave me the strength to be able to successfully complete this formative process of great importance in my life as a professional.

I thank to the professor José Tiberio Hernández, who accompanied me, guided and supported me in the Thesis process as my advisor. It was fundamental for me to have had his full support in this process, because thanks to him I learned about motivation, patience, to challenge me and about good practices to be an outstanding researcher. I thank God for having the opportunity to have had the support and guidance of him as my advisor.

Special thanks to Javier Peña, who collaborated me continuously in the process of my Research Thesis. Thanks to him I could have user testing spaces, in addition to always being attentive with such warmth, always being willing to collaborate. Similarly, I thank the research group "SUR" who voluntarily collaborated me with the study and all their valorous comments. To the professor Luis Ángel Guzman, thanks for the interest in the present study and for allowing me to have evaluation spaces with your undergraduate students. Without the participation of all these people, this project would not have culminated successfully.

1. Document's Structure

This document describes a problem and a context that focuses on the design, evaluation and use of tools in the search for optimization of the analysis processes and preliminary decision making in the stage of Exploration of data sources with representative characteristics of time and space, particularly. This begins by presenting the context that motivates the realization of the project, followed by its respective justification (chapter 2), along with the definition of the problem and the presentation of the objectives of the project, both general and specific. In chapter 3, the theoretical reference and state of art is elaborated, presenting previous projects and tools of interest for the study.

Chapter 4 continues, advancing with the presentation of the tool designed, in general and architectural terms, continuing with the Study Case, in Chapter 5 used through which was based the description and detail of the experimental phase in Chapter 0. Finally, the results obtained are presented complemented with their respective analysis (chapters 7 and 8), closing with the respective conclusions for this study in chapter 9 and the proposed future work proposal to give continuity to the project in chapter 10.

2. THE PROBLEM

Throughout this chapter, the problematic of interest for this research project is formulated, structured and delimited. In addition, the objectives to be achieved are presented, in addition to the respective justification to address the problem approached.

2.1. Problem Statement / Motivation

By its part, the analysis of spatiotemporal urban data has evolved along the last 30 years. Data heterogeneity and segmented analysis is a frequent issue in this context. As mentioned in the paper "*Big Data, Analytics and the Path From Insights to Value*", the smartest organizations are embedding analytics to transform information into insights and then actions, connecting somehow the performance and the competitive value of analytics (LaValle et al., 2011). Besides the latter, information must become easier to understand and act upon, depending on how the data is provided to any visual representation. Nowadays, data analytics is turned to be the future of management of various activities in the industry. As presented in the Figure 1, much more of the activities followed in an analysis done for Enterprise Global Executives in the U.S., requires analytics to fulfill their necessities on the decision-making process, more than their own intuition.

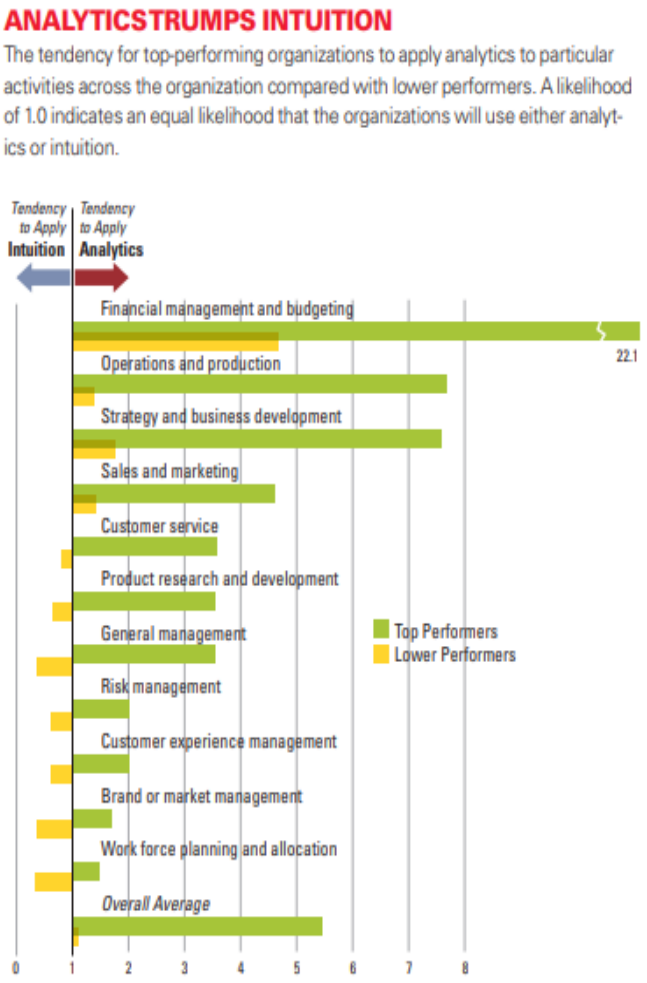


Figure 1 Analytics Trumps Intuition (LaValle et al., 2011)

Therefore, the effective and innovative use of data analysis tools has been a hallmark that distinguishes the most successful companies. It provides the necessary knowledge to make business decisions. For this reason, there are multiple tools on the market that allow users to know and analyze data. Known tools such as Tableau, Power BI or Google Analytics, among others, are well known examples for this, compared to traditional Business Intelligence.

Figure 2 Known tools for Data Analytics in the Market

Being the previously mentioned tools as an example of the current market are very efficient allowing the realization of numerous advanced visualizations with which to make the most of the data, decisions in selection of analysis tools, they often depend on the price and type of licensing. For example, for the cases of Power BI and Tableau, the first one is much more affordable, by costing Tableau seven times more approximately. However, we must consider all the factors mentioned, since it is useless to acquire a tool that fits the budgets but is not aligned with the needs and ends up falling into lack of use. Moreover, most of the tools have the “Freemium” services, which terminates in involving pricing when reach any point of analysis or needed greater features. This same happens with Google Analytics, that even if a web application which enables a bigger accessibility to the users as being in the Internet, has a pricing model, such as among other tools in the market.

On the other hand, barriers on becoming more data driven for organizations are based on managerial and cultural rather than related to data and technology. As can be appreciated in the Figure 3, even if the leading obstacle in terms of visual analytics’ adoption is the understanding on the alternatives of analysis to carry their needs (e.g.

lack of skills), as also costs and time in intern organizational terms, knowing the advantages or benefits of carrying out of this, though.

Figure 3 The Impediments to Becoming More Data Driven (LaValle et al., 2011)

Thus, through the development of this project, is intended to let the users to result in a customizable analysis to support the finding of new insights in a web approach unlicensed and currently without pricing for exploratory analysis terms as it is thought to be. Additionally, allowing the user, not only to visual query any view, but to explore and to analyze multi-scale attributes within the space and the time. With this web approach the alternatives of accessibility to data analytics will be fulfilled in such way that the users are permitted to find insights with such coordinated analysis within the multivariate, spatial and temporal attributes in a much more accessible manner.

Furthermore, by providing coordination features between views within the framework of the what, when and where paradigm, as described in subsection 2.2, the project

provides different analysis alternatives that imply various attributes, such as the temporal or the spatial ones. This study provides a palpable case study that seeks to familiarize users with the tool in graphical and analytical terms, thus gaining benefits in the discovery of new insights in terms of data exploration, by the fulfilling of challenges previously posed.

Then, if this project is put into practice, analysts from different perspectives and knowledge could make use of the tool developed in terms of support for decision-making processes, being considered as a facilitator seen from different analysis points of view. The latter, considering that the tool provides joint analysis features that allow the user to obtain graphic reports in a friendly and flexible manner for each of the users, regardless of the analysis variants provided by each one.

2.2. Justification

The realization of this research is of great interest to the entities that require an analysis of information with spatiotemporal particularity, because through this proposal the interaction between the main questions and / or components carried out in an analysis is achieved: (1) WHICH: Identification attributes, (2) WHEN: temporal attributes, (3) WHERE: spatial attributes and (4) WHAT: descriptive attributes.[6] In this way, summarize, compare, analyze, navigate, are the main tasks performed in this first preliminary tool resulting. With the above, it is possible to denote the importance of visual and / or graphic consultation, in order to involve more users who are not necessarily experts for an objective decision making.

The study significantly favors the analysis of large amounts of information, since it gives the user a sense of how the data is in terms of quality, distribution and outliers in a descriptive manner. These points, serve as input in terms of time and costs for the entities involved, providing ease in the understanding, comparison and analysis of a group of heterogeneous data in a short period of time, being all result of graphic interactions of interest.

2.3. Research Question

Considering the Problem Statement and the project's justification described above, the problem of interest is to favor the exploratory analysis of spatiotemporal data to let an analyst find insights in an effortless way than perhaps is possibly currently done. The proposal of this project consists in the design and the development of an interactive platform in a web approach, to ease a user, not necessarily an expert, to interact with its desired loaded data within synchronized views within the paradigm rounding the WHAT, WHEN and WHERE questions, reaching a Web environment approach for Visual Analytics.

This last one, captures the main question that is covered within the development and study of the present project. To the author's knowledge, a previous study that responds particularly to this problem is not present:

> *Does a web approach for multidimensional visual analytics through synchronized views, facilitate the analysis of spatiotemporal data in terms related to the paradigm of What, When and Where, eliminating accessibility, costs and licensing restrictions to users?*

2.4. Overall Objective

Support the finding of new "insights" from a specific visual analysis using interactive elements, in an objective and agile decision-making environment, based on the outline of Visual Information-Seeking Mantra of "Overview first, zoom and filter, then details-on-demand", taking in account the dimensions of space (Where), time (When), and multivariate content (What).

2.5. Specific Objectives

OE_1: Design, develop, detail, validate and document the technological tool for exploratory data analysis with spatiotemporal characteristics, orienting its preliminary design to be user-friendly and interesting.

OE_2: Elaborate and detail an experimental protocol that allows the measurement of variables involved in the validation of the tool serving as a facilitator for exploratory data analytics.

OE_3: Based on a use-case and a set of analysis task associated with, perform experiments in order to evaluate the proposed tool.

OE_4: Analyze the results of the entire project process, identifying advantages, limitations, problems and improvements compared to the proposal.

3. THEORETICAL REFERENCE

Throughout this chapter, will be described what has been worked lately about Data Exploration and Data Quality, principally, in general terms and as also regarding Spatiotemporality, specifically. That is, will be briefly described what is and how these works pose Data Quality, and for the Data Exploration towards Urban Planning, which are the types of analysis usually made.

3.1. Theoretical and Conceptual Framework

Reaching what Data Quality means correctly and following the preliminary conceptual framework mentioned in the work of Wang and Strong [4], the quality of the data refers to the accessibility, interpretability of the data, and the relevance and accuracy of it. That is, each category targets a set of dimensions towards believability, objectivity, completeness, traceability, the variety of data sources, value-added, timeliness, ease of operation, flexibility, ease of understanding, representational consistency and concise representation. Additionally, depending on the attributes or dimensions treated within a data source, four principal categories results: (1) Intrinsic, denoting that data have quality in their own right, (2) Contextual, highlighting the requirement that data quality must consider within the context of any specific task, (3) Representational and (4) Accessibility, emphasizing the importance of the role of systems. That is that high-quality data should be intrinsically good, contextually appropriate for the task, clearly represented, and accessible to the data consumer.

3.2. State of Art – Related Work

Towards the analysis and exploration of big amounts of data that is usually yearned, a visualization widget for summarizing, exploring and navigating multivariate datasets was implemented to achieve this current challenge. That is, Guerra introduced and evaluated the tool called "NAVIO", which displays full summaries, allows sorting and filtering on ranges and values, and keeps a visual trail of the queries, allowing users to navigate and explore the data effortlessly. As Guerra mentions in his paper [2], Summarization, referring to getting a general idea of the dataset as a whole, Navigation, and Exploration, involving the availability to run specific desired queries, are the three main tasks that Navio addresses.

It is stated how Navio displays a full summary of the dataset right from the start, displaying missing values, patterns, and distributions, giving the user a basic notion of the data completeness in an effortlessly way. Users then can explore the data while focusing on areas of interest by performing dynamic queries through in visualization set of selections, providing the trail of what the users perform. These set of characteristics are usually missing in the type of tools such as Tableau.

Additionally, to also support users without programming training, is presented “SHIPYARD” as a contribution where users can drag and drop their data and get a Navio visualization for understanding and exploring it, with the availability of setting up the variables or attributes involved by how the user desires. Notwithstanding these previous preliminary advantages, as running on a web browser as a characteristic of flexibility, both Shipyard and Navio gets to have low scalability as a result, in comparison with such tools as Tableau, regarding their limitation on the support of significant bigger datasets, usually over 400Mb.

Among a large group of commercial tools that support visual analysis, such as Infozoom or Tableau, the latter is the best known and common to achieve these visual tasks easily. Due to its nature, is the user who decides, among the dimensions or values, to visualize dragging and dropping each element on a canvas. As stated in Guerra’s work, this tool is great for the navigation, but it doesn’t lend easily to summarization or exploration. Being a common commercial tool, the tool is intuitive when you have knowledge of what you

are looking for, emphasizing its advantages over navigation tasks, supporting big amounts of data and processing them with a high efficiency.
Current works with variety of temporary query tools such as “TEMPEST”[9] allows the user to select arbitrary combinations of months, days and times of day and see what happened at this time. Furthermore, another type of interactive time filter in the “STRAD”[10] project where the user selects the period of time that he wishes to see by means of a slider where the consultation period is specified, placing the start and end date and time, allowing to visualize on the map the trajectories of ships that moved during that time.

Moreover, “TrapezoidBox”[8] is a reference tool for the spatial query, implemented on Google Maps, in which the spatial proximity queries are made by means of a trapezoid, where users can drag the four vertices to change the query condition and the result is seen in the map by the circular regions which represent the ranges satisfactory distance from the place of interest.

By its part, “ANIMAP”[15] with the application of MARS as simulation model, has as aim communicating simulation results to different stakeholder groups such as transport planners, traffic modellers and decision makers, to present information regarding the model design and model results at an adequate abstraction level. That is, chronologically spatial effects in transportation will be displayed to the users.

However, the comparison of spatial data in different instants of time aims to show the differences or proportions between the values for each moment and the values for the previous moment or at any time selected by the user.

For its part, Andrienko (2003)[7] proposes a topology based on the classification and evaluation of how such spatiotemporal data can help to resolve questions through exploration and the characteristics of the data they are applicable to. The proposal directly relates tasks to components of data, in terms of space (where), time (when) and objects

(what). That is, questions can be easily answered keeping only what users need to satisfy the query constraints, on what is called lookup and filtering.

On the other hand, Chen et al. (2018)[6] proposes a Visual Analyzer for Urban Data called "VAUD" which supports the visualization, querying, and exploration, allowing the Multi-source analysis by leveraging spatial-temporal and social inter-connectedness features, and selecting, filtering and aggregating across multiple data sources, which permits the extraction of information that would be hidden to a single data subset. Coincides with Andrienko when defining the resolution of a question on a query as the combination of four main component constraints, denoting the (1) which: Identification Attributes, (2) the when: Temporal attributes, (3) the where: Spatial attributes and (4) the what: Descriptive attributes. Moreover, is claimed the importance of visual querying, to engage more non-expert users.

In this order of ideas, even the on-the-fly queries and association of attributes are supported, VAUD requires users that must have a notion of networks, regarding the manipulation of nodes and the specifying of the conditions, to get to explore by zooming, panning and detailing the desired results.

Furthermore, Doraiswamy et al. (2018)[5] states that visual analytics systems such as their tool proposed as "URBANE", aim to empower domain experts to explore multiple data sets, at different time and space resolutions. In this proposal navigation and operations on map view such as panning, zooming, and rotating the view are accomplished through mouse interactions for analyzing multiple sources.

Speaking in terms of storage, querying and analysis both "VAUD" and "URBANE" softwares achieves a high efficiency. Different alternatives to this are thought considering each software's context and nature.

To finalize, will be briefly shown in the table below (Table 1) a brief summary of the couple of softwares mentioned above, in terms of (1) Storage, (2) Flexibility, (3) Accessibility, (4) Querying & Analysis, and (5) Main Spatiotemporal Interactions.

	VAUD ***Chen et al. (2018)[6]***	**URBANE** ***Doraiswamy et al. (2018)[5]***	**MARS meeting ANIMAP** **Emberger et al. (2012)[15]**
Storage	To enable cross-domain analysis by leveraging the spatiotemporal inter-connectedness, they build a sequence of STCs for spatiotemporal objects. The average memory consumption of an STC is 5Gb. Therefore, the total consumption for 22 STCs is about 110Gb. They store all STCs individually in the hard disk and construct a spatiotemporal index structure to accelerate the online query.	**Raster Join** Being a 3D Map proposal, a rasterization-based approach is thought to leverage current generation graphics hardware (GPUs), storing the different urban data sets in a 3D grid index of fixed size, where the dimensions correspond to the location (2 coordinates) and time.	**N/A**
Flexibility	**Multi-source Analysis / Interactivity** ● Manipulating Nodes: The	**Multi-source Analysis /Interactivity** Urbane generates queries for two	**Coupling of AniMap and MARS** A program written in VB-script converts MARS-

	analyst can create a node by moving a node onto the query view. ● Specifying Conditions: The analyst sets a query condition by first adding a node in the query view and then specifying the detailed conditions.	different operations—visualizing on the map and visualizing on the PCC [Exploratory View]. For both these cases, they execute Raster Join using a pre-configured 20 meter bound. However, users can change this bound if they require higher accuracy.	output text-files containing results of VENSIM-runs into javascript statements which in turn are interpreted by the AniMap-frontend. The provision of geographical data in the form of SVG-graphics is performed using ESRI's ArcView 3 automated through the using of several Avenue scripts. The spatial and temporal visualisation should be possible without access to the internet and without a (local) web server running on the computer where the presentation happens. The input data for MARS remain stored in Microsoft Excel (XLS) files. Does not support Multi-source inputs.
Accessibility	**NO** WEB	**NO** WEB	**WEB Application**
Querying & Analysis	**Multi-source Analysis**	**Multi-source Analysis**	**Control of time and the visible part of the region**: Time interaction bar

	* Exploring Results: The analyst is able to select one or more objects from the result node and place these in the scene view. The analyst can pan and zoom to explore details in the scene view. Furthermore, an analyst can explore detailed information by clicking an object.	The main goal of the data exploration view is to support the analyses of urban data at two different resolution levels—region and building. **Exploratory view** This visual representation is effective for analyzing multivariate data and can provide insights into the relationships between different indicators. Users can also filter regions by brushing the desired range of values on individual axes of the PCC. This updates the map by highlighting all regions that satisfy the filter constraints	Simulation smoothly or by jump by selection of Checkboxes. **Animated maps (VIEW-mode):** In AniMap/MARS only animated choropleth maps are used. **Comparing scenarios (COMPARE-mode):** Providing the direct comparison of the same output-variable generated by several MARS-simulation runs (which themselves are driven by distinct parameter settings).
	Space-time-cube based (STC) Constructed an STC for each time slice and	**Map View** This view is composed of a map	**Control of time and the visible part of the region:** The user controls the progression of time.

Main Spatiotemporal Interactions	uniformly subdivide the STC into a 3D grid for a given resolution, where the resolution is determined based on the analysis tasks. As such, a cell of the STC refers to a geographical location and a time interval in the time slice associated with the STC. Finally, we sequentially relate each record of each object into an STC cell by leveraging the timestamp and location information. The spatiotemporal data and associated STCs support fast querying of spatiotemporal information and facilitate indirect connections of objects by means of the spatial-temporal interconnectedness.	rendering component. The various menus and panels are overlaid on the map. Navigation and operations on map view such as panning, zooming, and rotating the view are accomplished through mouse interactions. The main menu allows users to control all the functionalities of the system. This includes loading or deleting urban data sets as well as polygonal regions that define the different resolutions. Users can then choose the data set to be visualized along with the visualization resolution.	**Map Viewing:** In VIEW-mode up to three variables (maps) can be drawn and animated simultaneously. Histograms underneath the maps pretend to visualize the frequency distribution of the underlying attributes (variables) at a certain time and are animated synchronously with the maps. **Comparing scenarios (COMPARE-mode),** offering the ability to compare outcomes of different simulation-runs of a MARS/VENSIM-model

		Multiple spatial aggregation queries can be generated based on the user interactions; thus, providing efficient support for these queries is crucial.	

Table 1 Comparative Analysis between Spatiotemporal Softwares

As can be briefly inferred from the summary above, some of the solutions in the same domain of S-PLOR-T considers a non-complete analysis for the paradigm through the WH* questions, which could be a first differentiation between them and the tool posed. That is, a limitation on the synchronized analysis between the paradigm's pyramid as presented in the Figure 4, is rarely met in current tools. On the other hand, for example towards the spatial analysis in a chronological manner, as ANIMAP's does, is usually thought to be linked to a map, which isn't the only way for it as S-PLOR-T briefly presents. However, each software shown above fulfills its own aim meeting functionalities that may not be able to meet with the first prototype of the tool posed, as described in Chapter 4, such as for example comparing features between scenarios, among perhaps others. For a further analysis about how S-PLOR-T is pretended to be a facilitator tool for analytics as a web approach in a spatiotemporal scope for this study, please refer to the Chapter below and later.

4. TOOL DESIGNED - Interactive Urban Spatiotemporal Data Exploration Tool

4.1. General Description of the Solution

A web application approach is proposed and shown within the exploration of urban data, with temporal and spatial features, that lets a user have a general notion of what is represented in the set analyzed. Everything will be turned around a

paradigm that expresses three main dimensions asked when having data sources with these characteristics. The elements WHAT, WHEN and WHERE will be describing the objects known as general/single attributes, temporal and spatial approaches, respectively. Therefore, as described in the [Figure 4] the relation between two “WH” question will help to answer the third remaining.

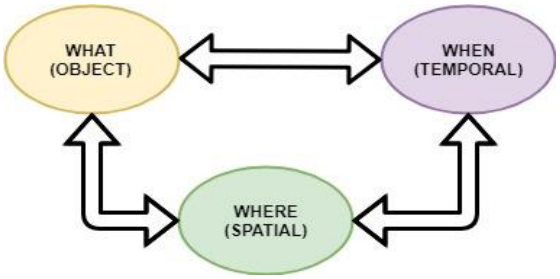


Figure 4 Paradigm Pyramid WH (WHAT / WHEN / WHERE)*

The visual solution proposed will be carrying different containers/modules that will be displaying linked views within the framework of the paradigm presented previously. Notwithstanding the main aim is to provide analysis around these three main dimensions, raw analysis over plain data, as a general statistical overview will also be carried.

Even if the first display will be proposing the user a configuration of each of the elements over a single screen (single web browser), the architecture that will be described below permits to have some flexibility over every physical container displayed. That is, the reorganization on the same multiple views over a single web browser, maximizing, minimizing or even closing them at the user's convenience, as well as setting a single or set of containers in multiple web browsers. Therefore, a user will have the opportunity to decide which of the views are important in its current analysis, or the opposite. Moreover, the link between every container will be explicitly or implicitly applied to what is called “cross-filtering”. That is, through the interaction over any module, adjusting over the other modules displayed will be made through linked filtering, even if the representation of the data is different between its views.

Now, the following presents and describes the general architecture proposed for the development of this first versioning of the prototype. Refer to Figure 5 presented and briefly described below:

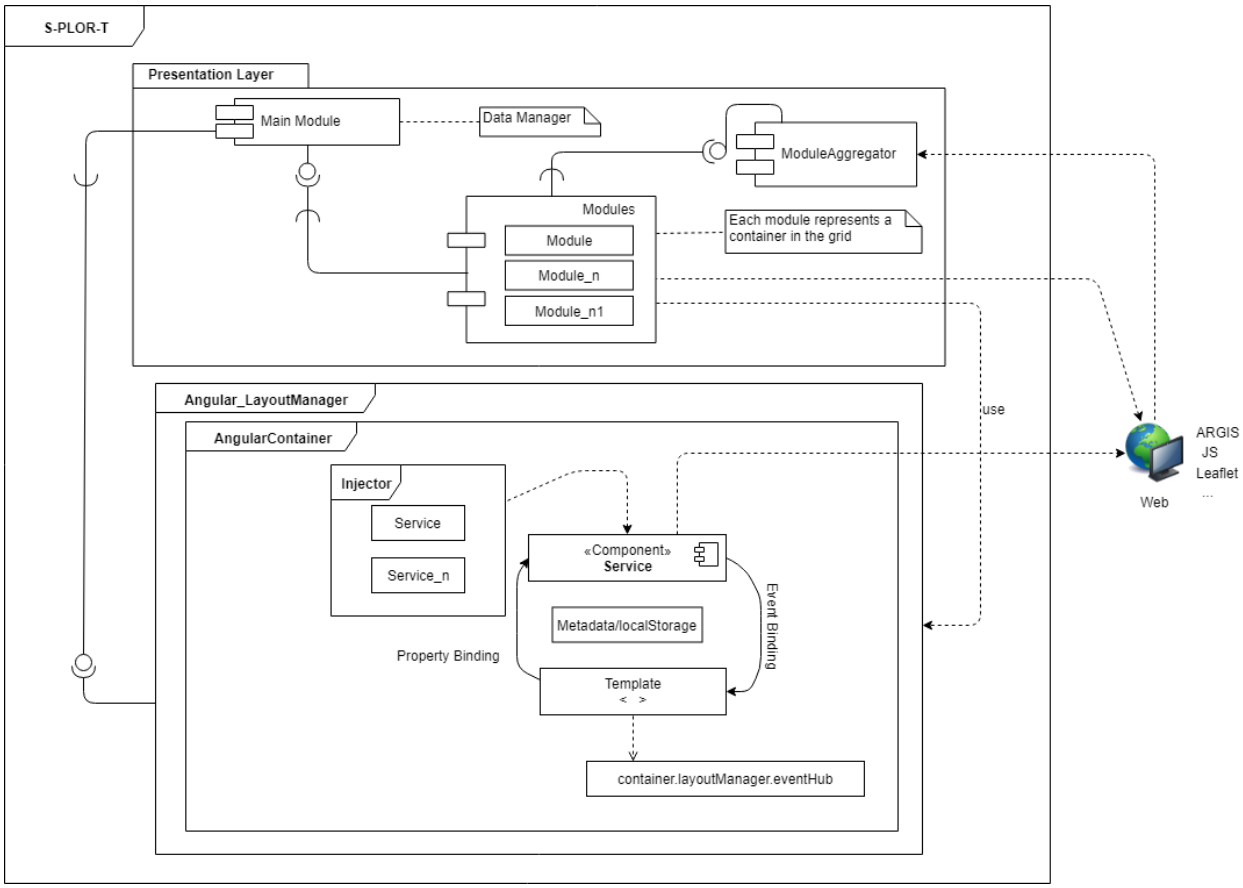


Figure 5 Architecture followed by the tool: S-PLOR-T (Component Diagram)

As can be appreciated above, on the first hand, the application consists of graphic components that depend on the interaction provided by the user, which is the involvement of logical components. That is, the elements presented in the presentation layer will depend on the communication carried out between the different modules that are displayed according to the discretion of each user. In this way, the presentation layer consists of "n" number of modules / containers that will represent information provided to the user (Summary / Statistics / Spatial-Temporal Analysis, among others), that can be added through an Aggregator Component that may or may not have web feed, at the user's need.

On the other hand, the main axis of communication lies in a parent module that is the Data Manager and which meets summary characteristics, which will depend on the interactions and the status of each of the modules displayed on the screen, by the current metadata or local storage per container. In the same way, all the modules represent a status update based on the interactions carried out in the other displays on the screen. The above, because all modules handle communication events through injected services per container. In this way, through a design manager and its event hub, the modules listen and receive constantly to the data provided by the data manager of the presentation layer.

Now, in a more detailed way about the data managing in a logical manner, from the diagram's layer below the Presentation's, can be seen three well differentiated parts, the components (component), the templates (template) and the services (service). Thus, the components are blocks of the user interface with specific functionality, writing the business logic and composing each other to form the view of the general application. The services will be everything that connects the components with something for the exchange of data per module. In this way, the components use the services to obtain the data to be presented in the templates.

4.2. Implementation environment and restrictions considered

The implementation will be carried out mainly using the D3.js library, Vega-lite, jQuery DataTables, Leaflet, AngularJS and JavaScript and other libraries that will mainly help to meet the expected use characteristics, such as the flexibility granted to the user in visual terms.

4.3. Stages of development

- Determination and application of the flexibility in the selection and display of visualizations before selections of interest made by the user in any of the analysis components: WHAT / WHEN / WHERE.

- Definition and execution of strategies to support additional information in spatiotemporal urban data sources in the prototype solution.
- Definition and execution of technical tests and user tests for validation of results, following the definition of a brief known study case.

4.4. User Interaction

The general interaction all over the whole proposed web application is described in this subsection. When first run the application's main server, it is displayed with its very first proposed configuration of the preliminary modules shown, as presented in the Figure 6.

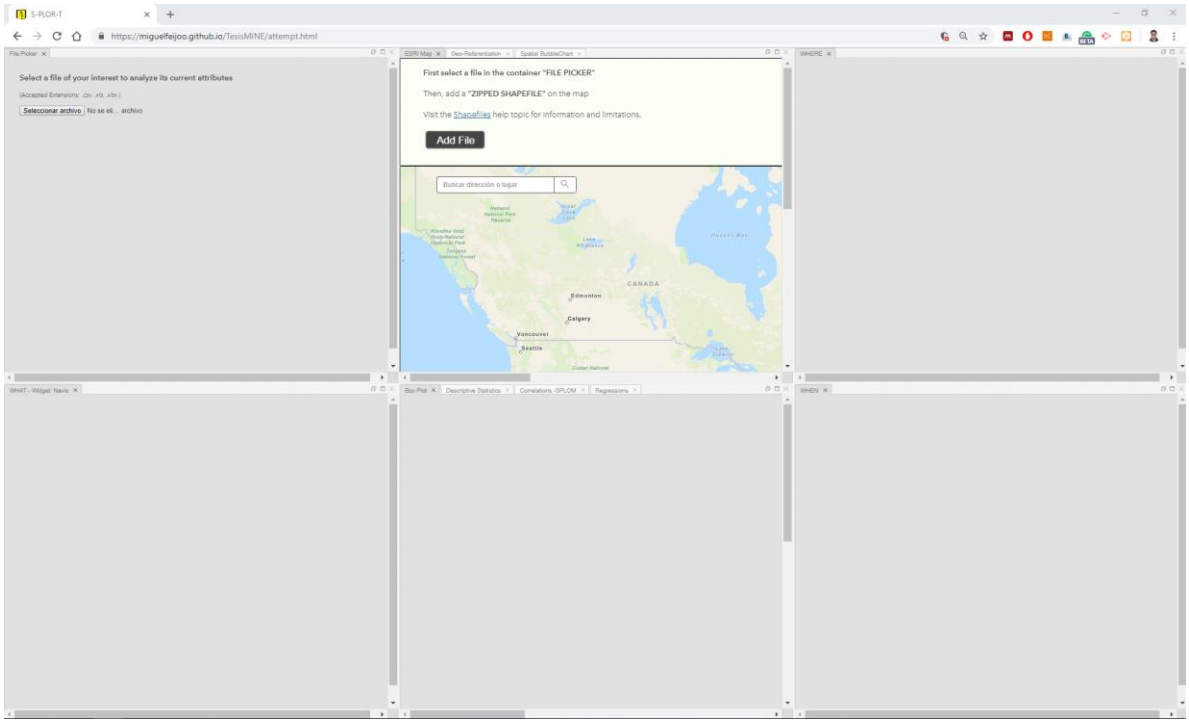


Figure 6 First View of the tool: S-PLOR-T

On the very first view, preliminary are represented six main containers, which are identified in the tab-names: (1) File Selector, (2) WHAT - Widget: Navio, (3) General Statistics, (4) Geo-Referentiation, (5) WHEN and (6) WHERE. All the views have the same data displayed. That is, any interaction over each view will be affecting the content displayed on the other views.

On the other hand, the user is able to load a data source that must contain spatiotemporal features. To do that, in the very first container at the upper left, as

proposed in the default configuration, the user only has to select a local *csv* file, as a preliminary extension restriction, as is represented in the Figure 7.

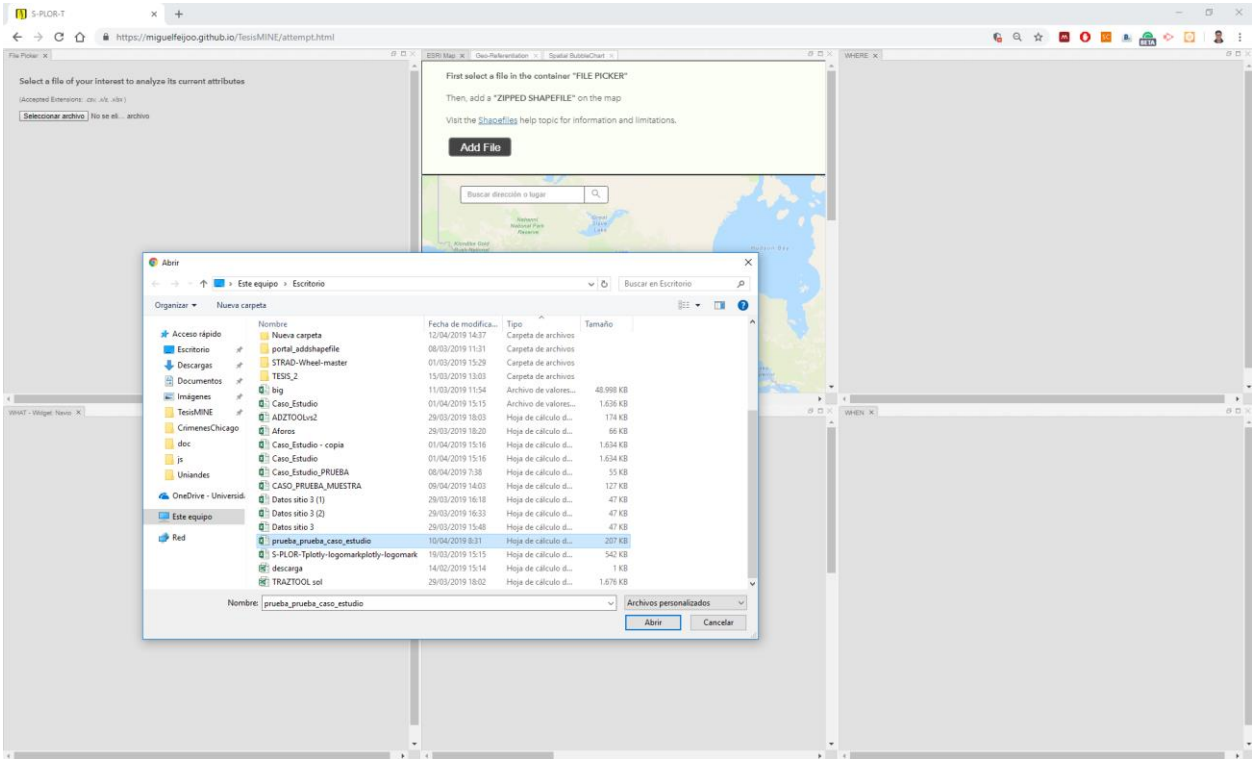


Figure 7 File Picker view – Uploading a file

As the user select the file to analyze, the data will be loaded to the remaining containers and each will display a first view around the statistical and the “WH” paradigm approaches, as the shows Figure 8.

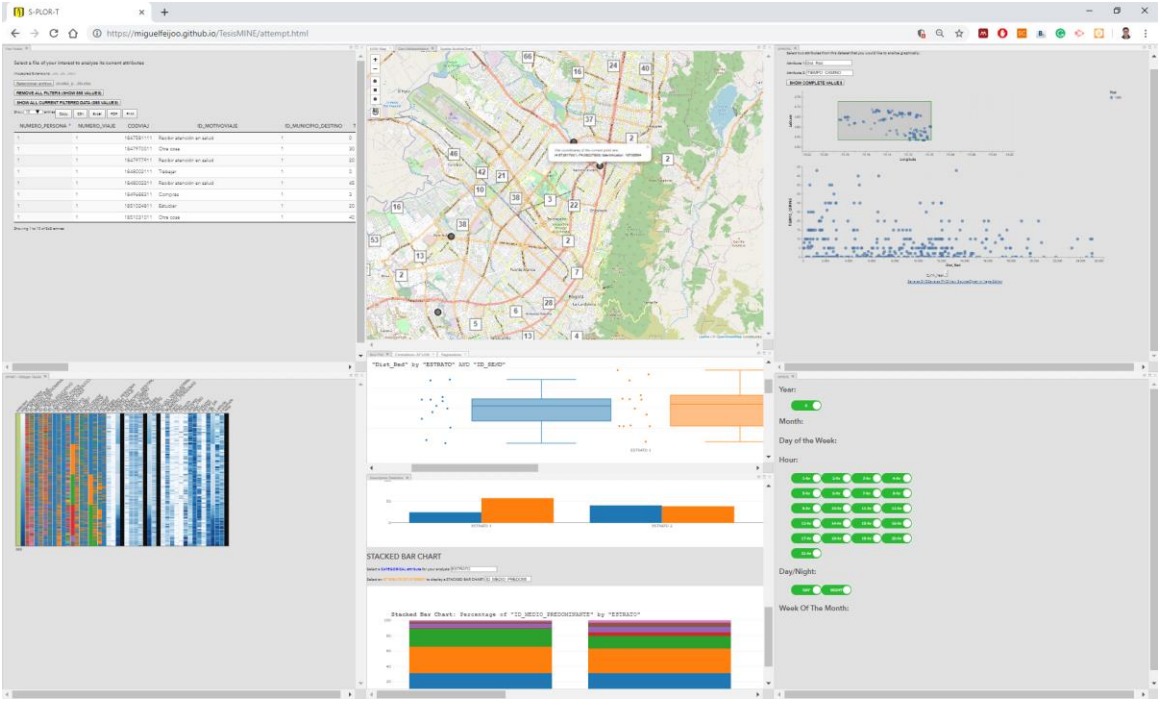


Figure 8 Charged general view of the tool: S-PLOR-T

As mentioned above, the user is free to re-configure as its own convenience any of the containers presented preliminary. That is, by a sustained click over each separate tab of any container, the user is able to move the whole module and its views to where is desired in the same single screen (web-browser), as is represented in the upper Figure 9 shown below. Moreover, the grouping feature at all the containers as tabs is also achieved by holding the click over any module and locate it in the upper tab section over another container, as represented in the bottom Figure 9. Additionally, every container has at its upper right a brief interactive bar, presented in the Figure 10, which let the user close, maximize/minimize or “pop-up” the container in another window, respectively from right to left.

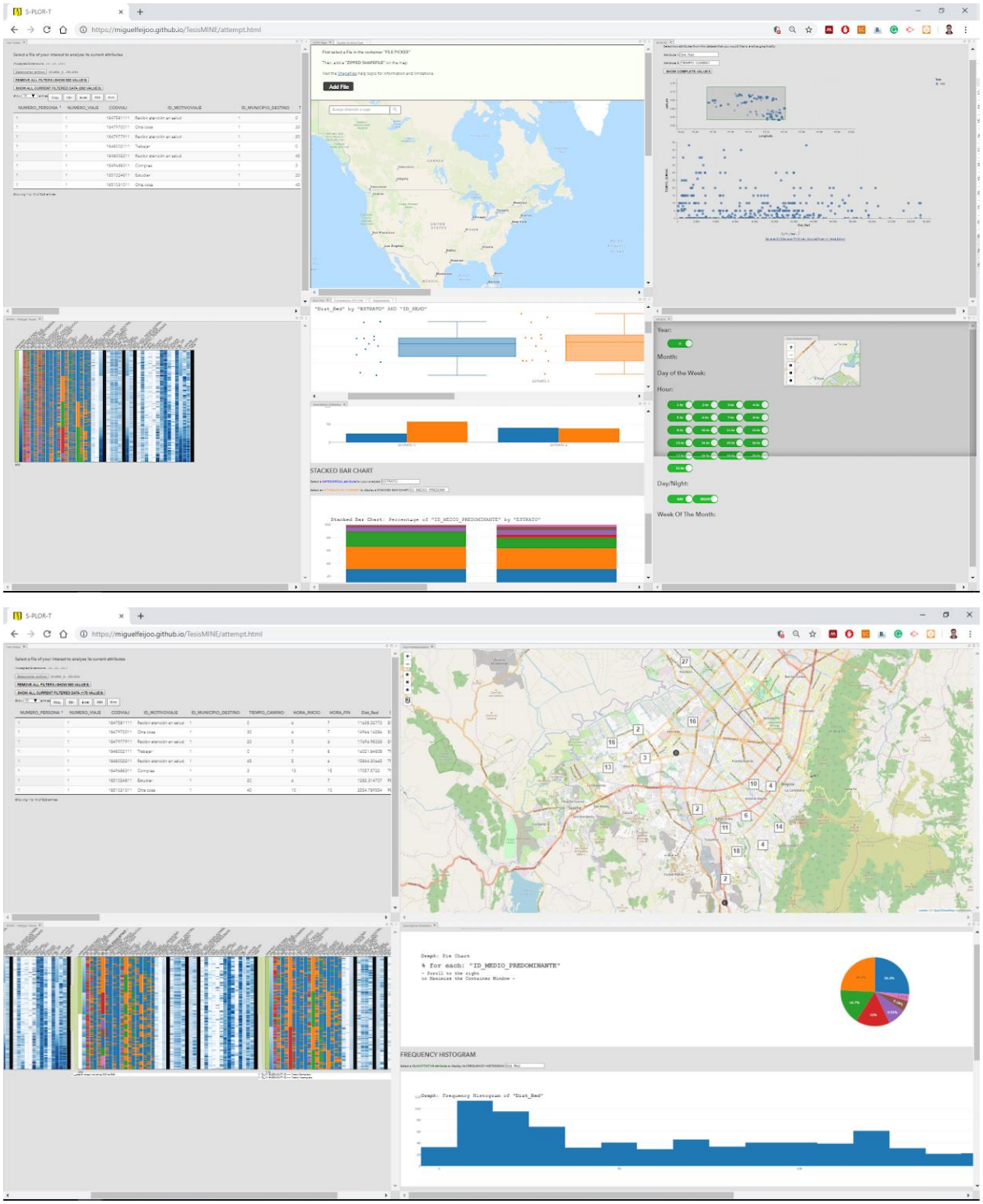


Figure 9 Interactions over and communication between modules/containers/views

Figure 10 Interactive bar by Container

Even when popping up any container in a new window, all the content that is displayed in it will be normally linked with the other open containers whether they are also popped up in new windows or not, through dynamic cross-filtering (through a bidirectional communication). Furthermore, the current state of each container will always by managed no matter the configuration is done over a single screen or multiple windows. That is, the state will let the user know the current analysis step and the approximate trace of the interactions carried out to achieve current results.

The definition and interaction over each separate container will be presented in a detailed way in the subsection 4.5 as also in a much more detailed way in Annex 12.1: *Usability Tool Guide*.

4.5. Detailed Specification

In this subsection, a general specification of the web application proposed, is presented in terms of the architecture and the form and interaction approach for each of the different modules.

For a greater understanding of the context of the architecture, the "module or container" and "service" concepts are treated as a means of communication between them. As previously mentioned in this chapter, an interaction between modules represented in a preliminary grid by default is proposed, which has the flexibility desired by the user. This feature allows the organization at the user's interest of all the different components displayed in a single screen (said in a browser, with restriction to a single "view" not expanded), as well as separating them in several windows, maintaining the common interaction between the different modules or applications, on what is called “crossfiltering”. Similarly, flexibility in the sense that the user can decide which views are preferred available for the analysis that is kept in mind, which implies that closing, maximizing, minimizing, grouping into tabs can be applied for each of the modules that contain the views of interest.

In the Figure 11 is represented a summary of how the architecture of the main application is worked. Basically, when the data source is first loaded to the server, a preprocessing is held in managing location and temporal features, particularly, in order to obtain a natural formal and common language through the whole application, no matter what data source the user is working with, and to obtain additional information that is not presented explicitly in the current data, such as new attributes as *"isWeekDay, day/night, dayOfTheWeek* and *weekOfTheMonth"*.

Using $scopes, AngularJS permits a first bidirectional communication between a Service and any Module. That is, by setting a general $scope of the data, the very first module can obtain all the data loaded of a single source. Once the module gets the whole data objects (array of JSON objects), through events that are held by EventHubs of a container managed by a LayoutManager, with *container.layoutManager.eventHub.emit(callbackFunction)* and *container.layoutManager.eventHub.on(callbackFunction)*, every container opened in a single window or in multiple windows are constantly communicated and updated whenever happens any change (such as filtering). The *.emit()* method will always be sending current data from any module to the others, which are always listening and receiving that same state of the data with the *.on()* method.

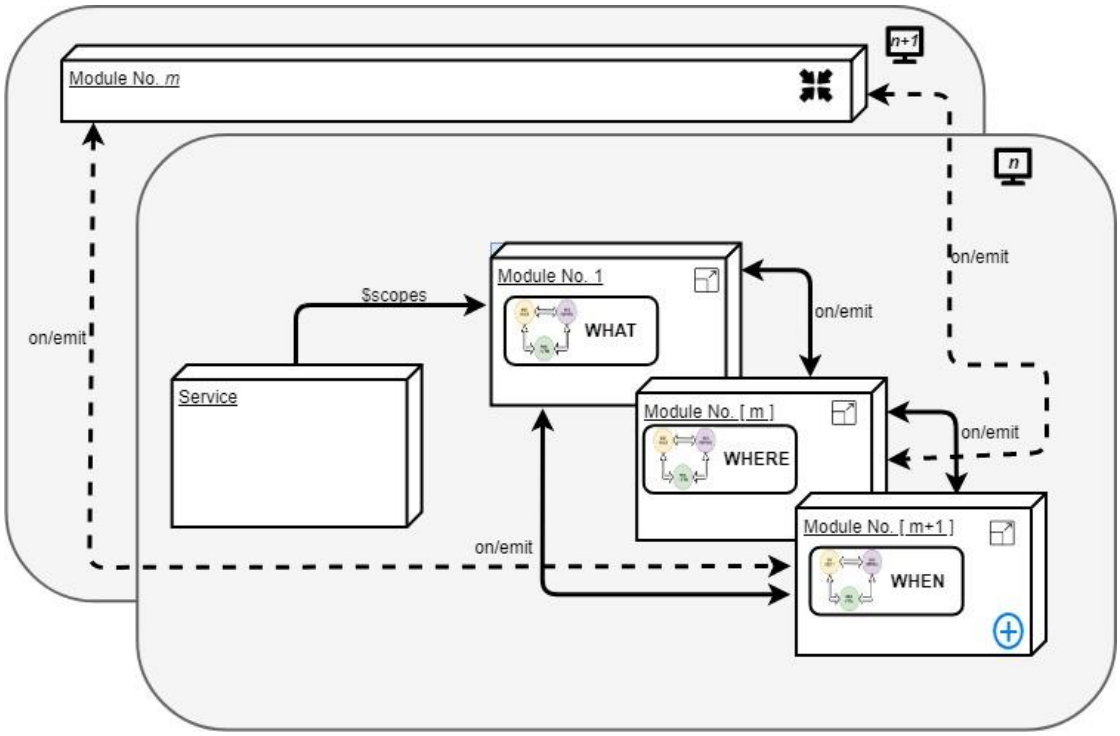
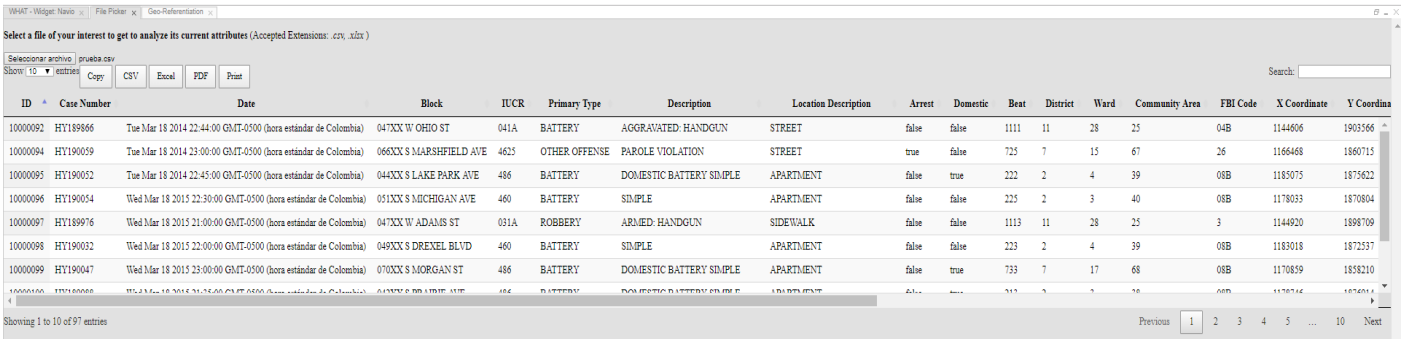


Figure 11 General Communication/Interaction between modules

Now then, it is shown below the detailed specification for each view as a descriptive model. That is, the detailed definition and interaction of every separate module presented preliminary, answering in this way WHAT and HOW questions, particularly. For this, check the subsections 4.5.1 to 4.5.6, as it follows.

4.5.1. File Selector Module

This module is thought to be the main data loader, which permits the user to select the file source as desired. Particularly, the user is restricted to select files whose path has extension *.csv or .xlsx*, as can be seen in the Figure 12. When selected the desired file, it is loaded and reported in a dynamic table, which has features as pagination, show n numbers of entries per page, filter by text (in the search input space), order ascending/descending by any of the attributes loaded, and determine if needed or wanted, to copy on clipboard, print or export as Excel, CSV or PDF the current visible data displayed on this table.



ID	Case Number	Date	Block	IUCR	Primary Type	Description	Location Description	Arrest	Domestic	Beat	District	Ward	Community Area	FBI Code	X Coordinate	Y Coordina
10000092	HY189866	Tue Mar 18 2014 22:44:00 GMT-0500 (hora estándar de Colombia)	047XX W OHIO ST	041A	BATTERY	AGGRAVATED: HANDGUN	STREET	false	false	1111	11	28	25	04B	1144606	1903566
10000094	HY190059	Tue Mar 18 2014 23:00:00 GMT-0500 (hora estándar de Colombia)	066XX S MARSHFIELD AVE	4625	OTHER OFFENSE	PAROLE VIOLATION	STREET	true	false	725	7	15	67	26	1166468	1860715
10000095	HY190052	Tue Mar 18 2014 22:45:00 GMT-0500 (hora estándar de Colombia)	044XX S LAKE PARK AVE	486	BATTERY	DOMESTIC BATTERY SIMPLE	APARTMENT	false	true	222	2	4	39	08B	1185075	1875622
10000096	HY190054	Wed Mar 18 2015 22:30:00 GMT-0500 (hora estándar de Colombia)	051XX S MICHIGAN AVE	460	BATTERY	SIMPLE	APARTMENT	false	false	225	2	3	40	08B	1178033	1870804
10000097	HY189976	Wed Mar 18 2015 21:00:00 GMT-0500 (hora estándar de Colombia)	047XX W ADAMS ST	031A	ROBBERY	ARMED: HANDGUN	SIDEWALK	false	false	1113	11	28	25	3	1144920	1898709
10000098	HY190032	Wed Mar 18 2015 22:00:00 GMT-0500 (hora estándar de Colombia)	049XX S DREXEL BLVD	460	BATTERY	SIMPLE	APARTMENT	false	false	223	2	4	39	08B	1183018	1872537
10000099	HY190047	Wed Mar 18 2015 23:00:00 GMT-0500 (hora estándar de Colombia)	070XX S MORGAN ST	486	BATTERY	DOMESTIC BATTERY SIMPLE	APARTMENT	false	true	733	7	17	68	08B	1170859	1858210

Figure 12 View of the summary of the file picking container

4.5.2. WHAT - Widget: *Navio*

Within the main paradigm "W's (WHAT-WHERE-WHEN)" this module answers particularly the first dimension, mainly focused on the attribute feature as itself of the loaded data. As represented in the Figure 13, this module manages all the data loaded in a graphic way to let the user to have a notion of how it is set. *Navio* displays a full summary of the dataset right from the start, displaying missing values, patterns, and distributions, giving the user a basic notion of the data completeness in a effortlessly way.

In the interactive approach, users can explore the data while focusing on areas of interest by performing dynamic queries through in visualization set of selections, providing the trail of what the users perform, as the figure below represents. In order to achieve the visual query, the user has to hold a left click over the widget and select the desired data over brushing. On the other hand, ordering and filter by single attributes is also permitted in this visual analysis.

The data displayed in each visual set, is the one the other modules will keep listening constantly (called as Visible Data). That is, whenever any selection over the widget is done, the data displayed over the other views will also keep changing, always to the visible one.

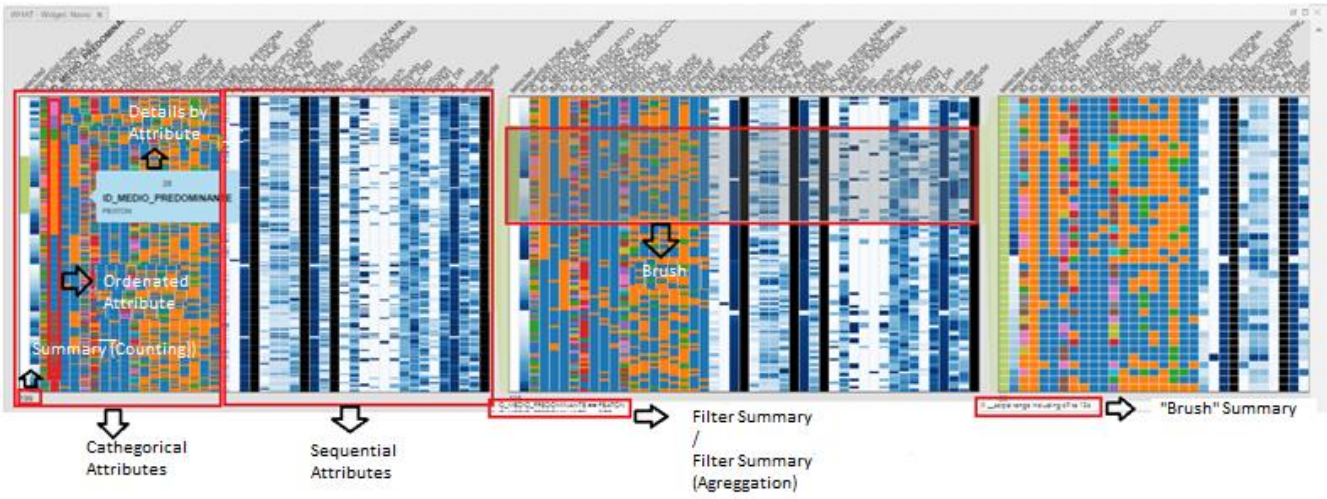


Figure 13 Navio's widget description and its main interactions

4.5.3. General Statistics

As a general approach, this module lets the user have a notion of the simple statistics of a single attribute by selection at its own convenience. As is represented in the Figure 14, a general boxplot is displayed with the distribution of the values of the attribute selected. Per year is shown a representation of the quantiles calculated, and with it the distribution that fits between different levels of the boxplot or the whiskers, as also information that is higher or lower to the maximum and minimum data, respectively.

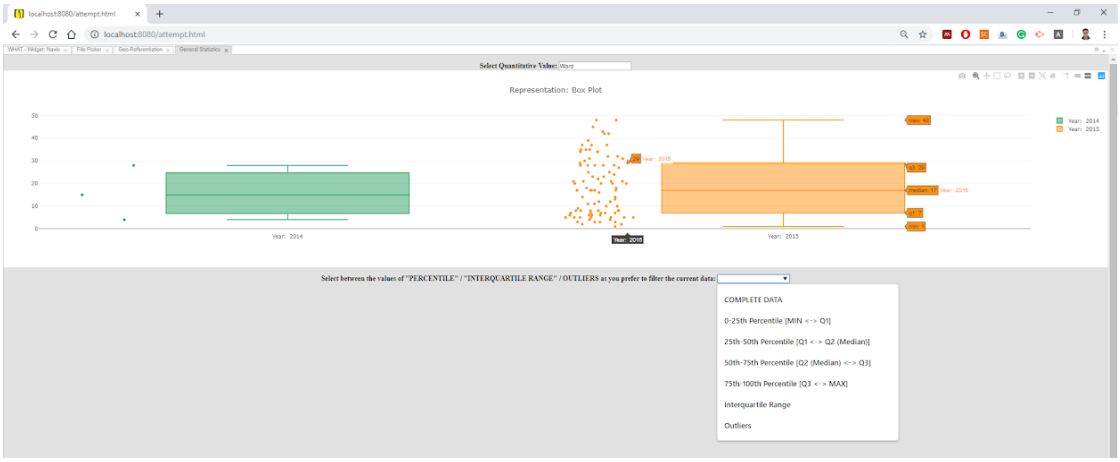


Figure 14 Statistics container: Box Plot Analysis

In the interactive approach of this single view, the user is able to see the current distribution and its boxplot (from the visible data displayed on the widget Navio), with its features (eg. single value). Furthermore, a selection over the legend permits the filtering over the preliminary graph. To return to the complete representation, the user has to select over the legend attributes to hide/unhide values in the graph. Moreover, as a zooming feature, the user is able to select by brushing the desired data to have a better notion of what is initially displayed. To return to the general view, the user has to double click over the graph.

Additionally, the user is also able to select over the dropdown below the graph, to filter over the current and other modules the actual visible data, considering different levels over a single box and whiskers representation, as follows:

- *Complete Data*: Raw loaded data.
- *0-25th Percentile [MIN <-> Q1]*: Range that represents all values between the lower whisker (minimum value) and the first quartile.
- *25th-50th Percentile [Q1 <-> Q2 (Median)]*: Range that represents the 25% of data that will be less than 25th percentile; 75% of data will be more than 25th percentile.
- *50th-75th Percentile [Q2 (Median) <-> Q3]*: Range that represents the 50% of data will be less than 50th percentile; 50% of data will be more than 50th percentile.
- *75th-100th Percentile [Q3 <-> MAX]*: Range that represents the 75% of data will be less than 75th percentile; 25% of data will be more than 75th percentile.
- *Interquartile Range*: Range that represents all values between the first and the third quartile.
- *Outliers*: Values that between all calculation of higher and lower whiskers, are bigger and lower than each, respectively.

On the other hand, other alternatives to increase the statistical analysis is provided by S-PLOR-T, such as graphics which let the user to get information from descriptive statistics, as shown in the Figure 15 below:

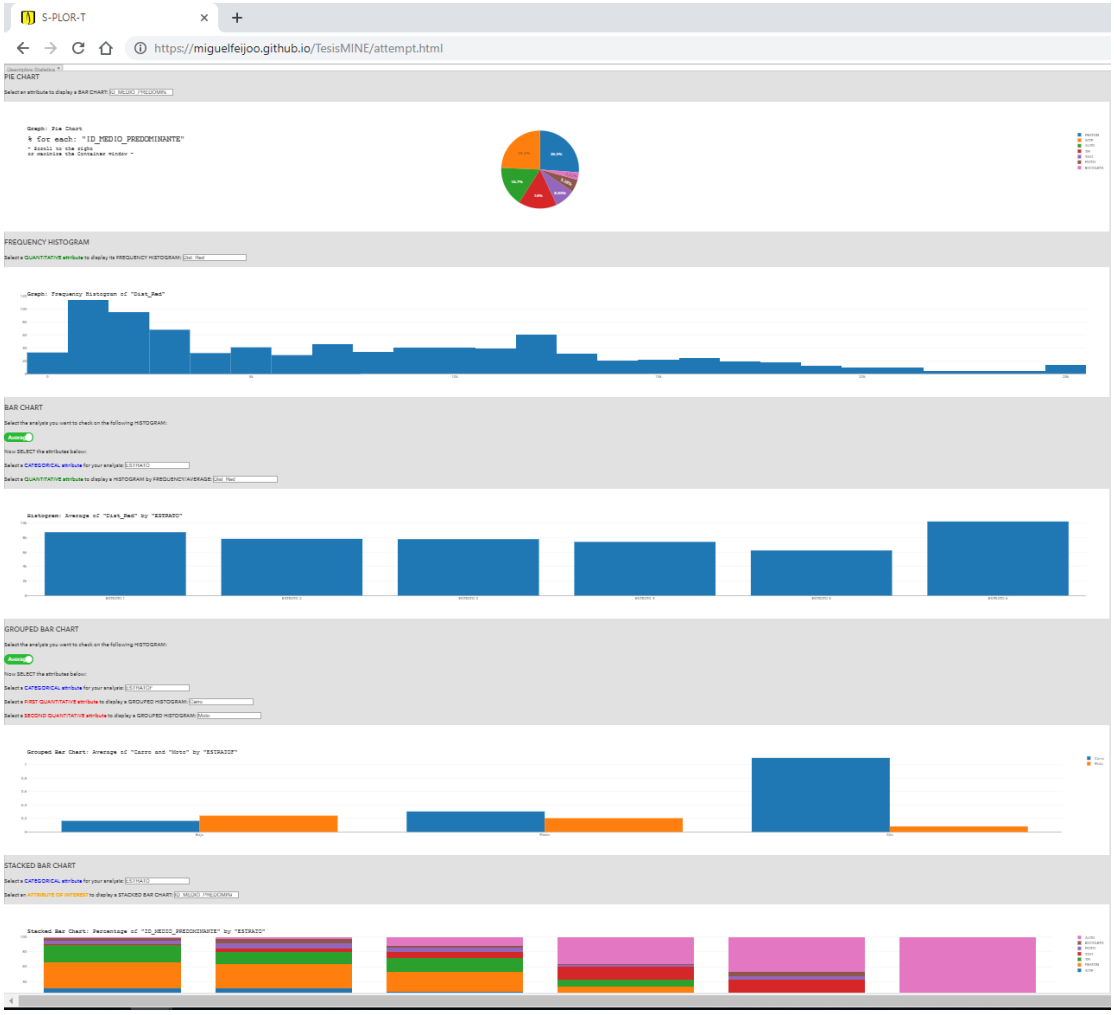


Figure 15 Statistics container: Descriptive Analysis

Moreover, correlational analysis between various attributes and between a pair of them can be also done as other alternatives in the analysis process, if needed.

For much more detailed information, please go to the Annex Section 12.1: "*Usability Tool Guide*". Important to say that you will find a guide in Spanish, due to the participants of the usability evaluation proposed and presented later, are all Spanish-speakers.

4.5.4. Geo-Referentiation

Around the need for understanding visually the spatial dimension presented within the loaded data, this module represents the set of coordinates obtained by the processing of Latitude and Longitude per Identificator (ID - “Key”) per year (if containing temporal features). That is, as can be appreciated in the Figure 16, for each year is a distinct color representation of the information displayed on a map (implemented following the library of “OpenStreetMap / Leaflet”). By clicking over any point of interest, the user is able to get the current registered coordinate, with its key-Identificator and, if existing, the Date linked with the “event”.

Not less important, whenever any additional module has an interaction (such as filtering by area), the information displayed over the main map will also be updated. This last will permit the user to have a zooming notion of what the analysis requires.

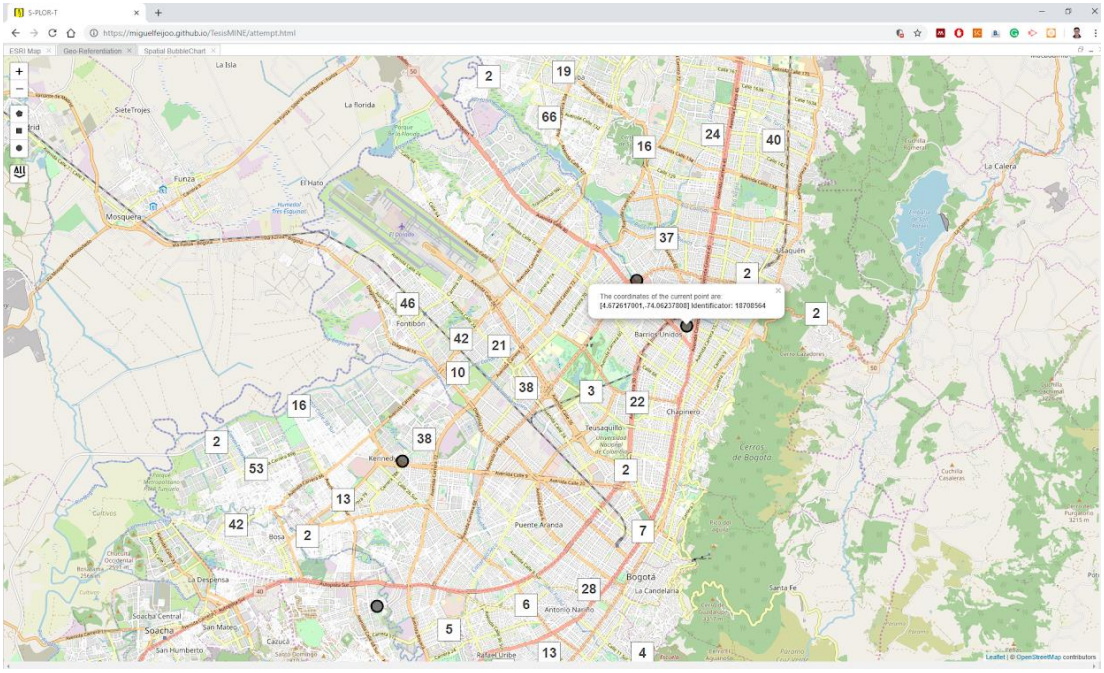


Figure 16 Spatial Container: Georeferentiation view

On the other hand, if needed and provided in the data, you can load and aggregate $ESRI$ information in terms of shapes or $geoJSON$ formats, providing additional interactions over the data, as shown in the Figure 17 below.

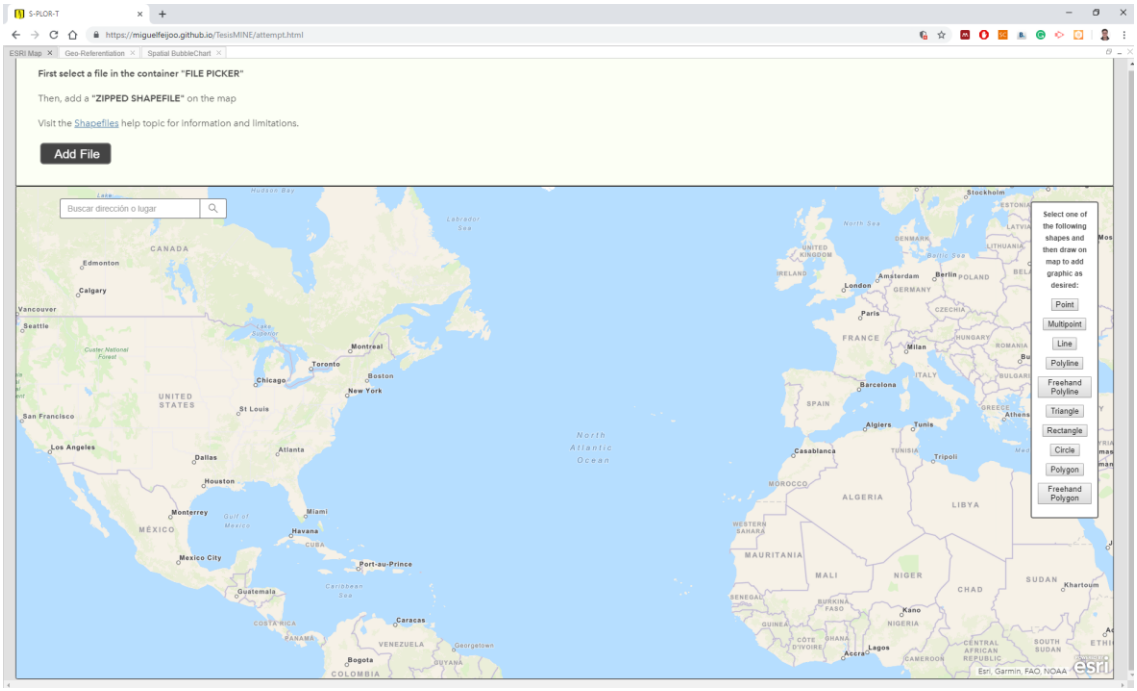


Figure 17 Spatial Container: ARGIS map view.

Moreover, non-map analysis, within the spatial component is also provided. The next subsection manages this in a better detailed way.

For more detailed information, please go to the Annex Section 12.1: "*Usability Tool Guide*". Important to say that you will find a guide in Spanish, due to the participants of the usability evaluation proposed and presented later, are all Spanish-speakers.

4.5.5. WHERE

Within the main paradigm "W's (WHAT-WHERE-WHEN)" this module answers particularly the second dimension, mainly focused on the spatial feature of the loaded data. By selecting any two variables of interest, the application will propose a graph that fits on the need of visualization. That is, between the combination of categorical and sequential attributes, following

the best practices that state how to reduce the cognitive effort of the user, the module will display a good approximation of visualization that could help the user to analyze attributes within the selections at its own convenience.

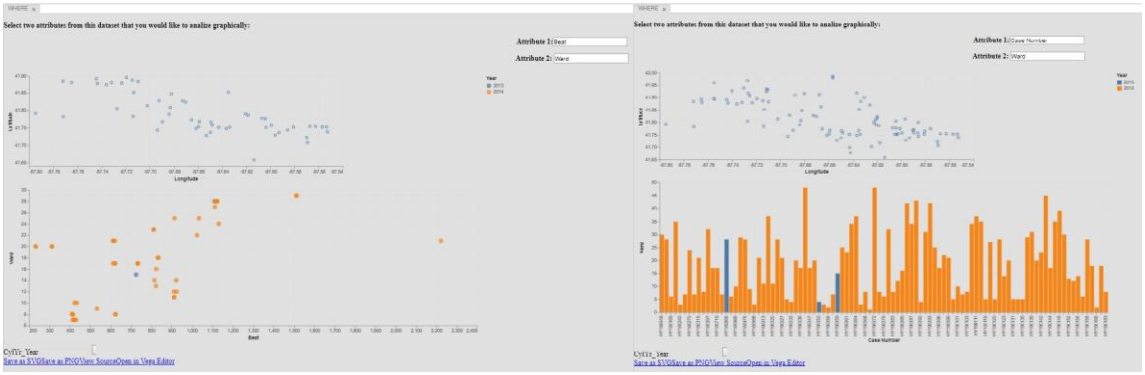


Figure 18 Spatial Container: Un-map analysis

On the other hand, besides getting to know the general static data of the currently chosen attributes, the user is also able to select by brushing over the spatial graph (presented above between the two graphs) which represents the coordinates in the plane (Latitude and Longitude), also presented visually in the georeferentiation map. By holding a click over the graph (scatter plot) and brushing, the information of the graph below will be also filtered. To return to the general view, the user has to click outside the brushing area previously drawn. This is represented in the Figure 19, which can be found below.

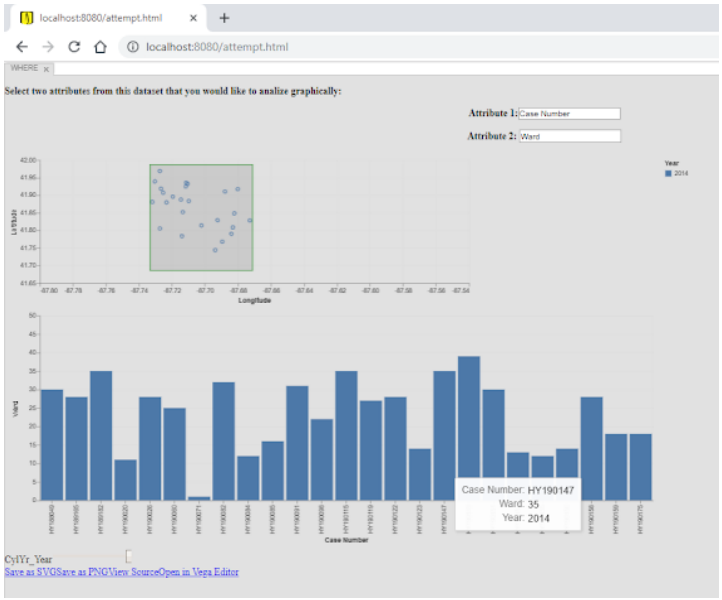


Figure 19 Spatial Container: Un-map analysis

As vega-lite provides, every set of visualizations can be saved as SVG or PNG by just clicking the links below both graphs, as the user desire.
Not less important, all the information displayed in this module is the current visible one over the widget NAVIO, representing the WHAT approach of the paradigm studied.

Additionally, other spatial non-map analysis can be held, as proposed in the prototype, such as Bubble Charts. Other alternatives can be also considered to be added to a second version of the tool.

For more detailed information, please go to the Annex Section 12.1: "*Usability Tool Guide*". Important to say that you will find a guide in Spanish, due to the participants of the usability evaluation proposed and presented later, are all spanish-speakers.

4.5.6. WHEN

Within the main paradigm "W's (WHAT-WHERE-WHEN)" this module answers particularly the third dimension, mainly focused on the temporal feature of the loaded data. That is, different characteristics that can be obtained from temporal attributes such as: Years, Hours, Months, Days, and so on. Within this container will be displayed additional information obtained from the temporal features that the loaded data may have, as possible.
Within this third dimension, represented as the Figure 20 shows, users will be able to filter temporal data as desired in terms of different levels of temporal granularity, by switching on and off any of the toggles as shown below:

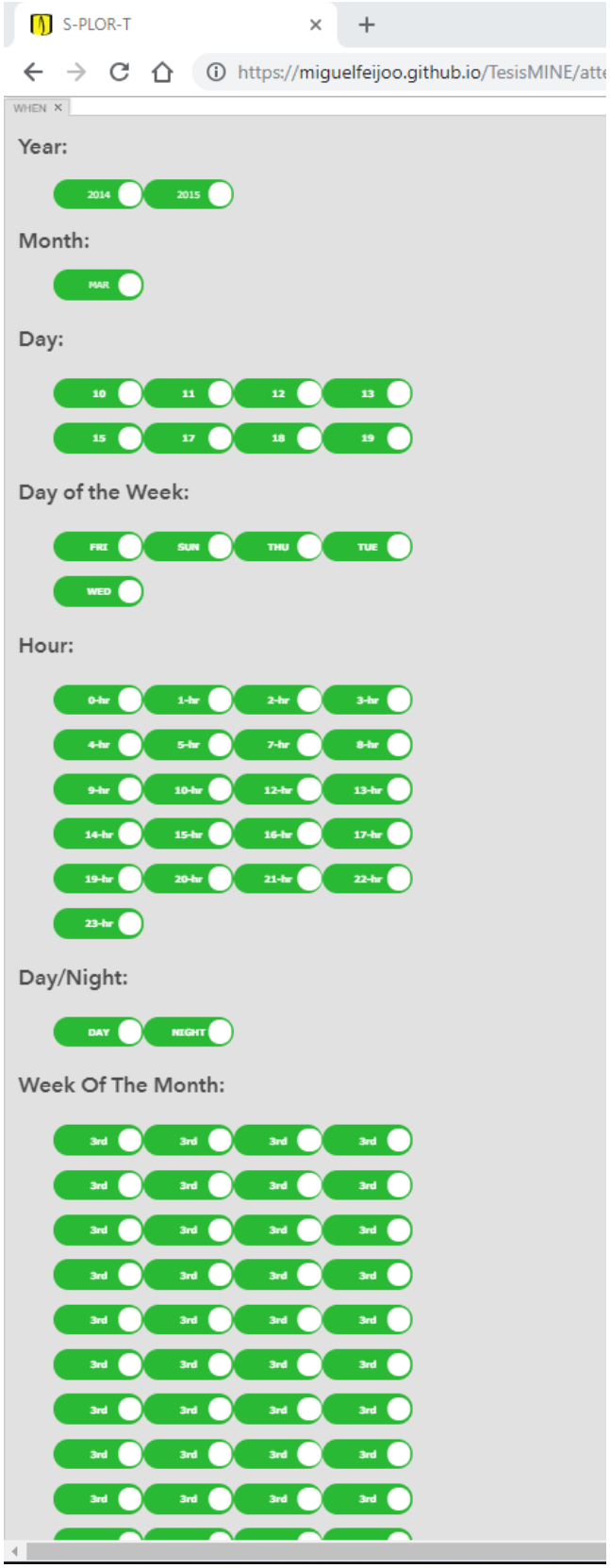


Figure 20 Temporal Container

For more detailed information, please go to the Annex Section 12.1: "*Usability Tool Guide*". Important to say that you will find a guide in Spanish, due to the participants of the usability evaluation proposed and presented later, are all Spanish-speakers.

5. STUDY CASE- DESCRIPTION: "Behavior of daily urban trips by home - Study case in Bogotá city"

5.1. Description and Motivation

Within the framework of the master's thesis in Civil Engineering developed by the researcher Javier Rodrigo Peña Bastidas, from the Universidad de los Andes, it is proposed as a brief case study, for this particular project to use a sample of a set of provided data for the investigation. Group: "Group of studies on urban and regional sustainability - SUR", which is a working research group of the Universidad de los Andes, of the academic unit: Department of Civil and Environmental Engineering.
The data provided is the data from the Household Travel Survey of Bogotá for 2015, which states: "... the most complete official source of information on mobility and travel patterns ...". This survey is the compilation of the one-day travel diary of people over 5 years of age living in participating households located in the area that includes the city of Bogotá and its 17 neighboring municipalities. This provides trip information of two points (origin-destination) per member per household. In addition, sociodemographic information is also included, such as age, socioeconomic stratum, means of transport, gender, etc. As detailed, the survey of this year (2015) compared to 2011, included more detailed characterization of the trips on foot, since it recorded all trips of more than 3 minutes since the previous exercises did not take into account for trips of less than 15 minutes.

The analysis of cities such as Bogotá, city in Colombia, as in terms of urban mobility, allows finding patterns that allow determining the level of development of the city in

terms of transport, considering the nature of distribution of households in the urban space. As a result, housing settlements are common in areas with low economic opportunities and low mixed land uses, generally in the periphery of cities caused by the fast-growing and uncontrolled nature of the population, together with the lack of Public Transport, leads to long travel times and distances. Economic and social conditions end up being fundamental in the gaps that currently affect the city.

Thus, getting to know those patterns of mobility of great amount of people in the city, may tell Bogotá's planners to make quick decisions just by obtaining descriptive analysis, as spatial and temporal analysis, in an effortless but objective manner.

Notwithstanding the analyst could have different alternatives of visual analysis, every separate question could get him time-spending, such as needing desktop apps, licenses, or just knowing how to code. With S-PLOR-T that will not happen as being a web approach for data analysis in an exploratory phase with synchronized views relating any question within the paradigm's questions: What, When and Where.

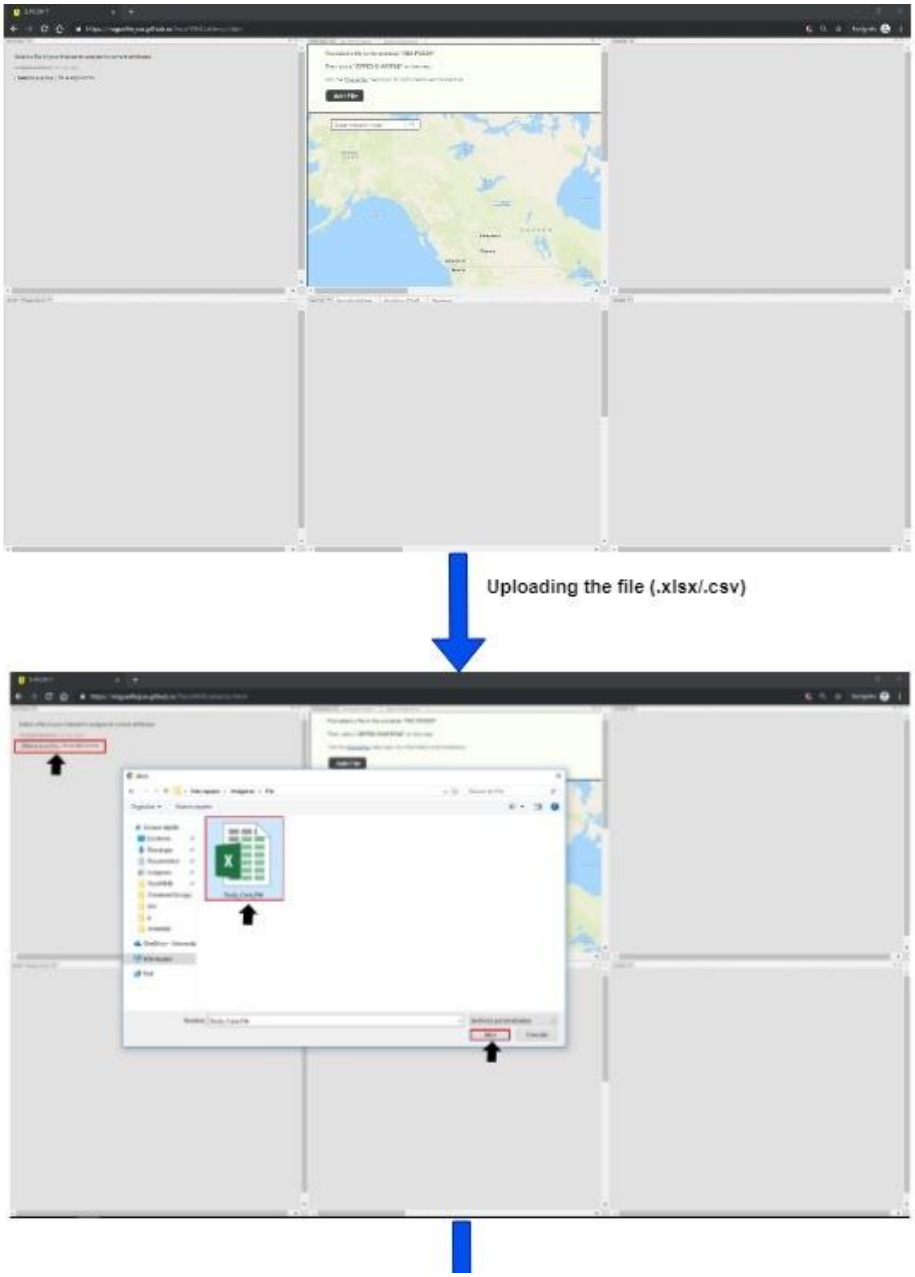
5.2. Storyboard - Process of Interaction

In this subsection, will be shown an alternative of process of analysis in graphic terms to situations that can be presented, considering the description of the use case of the previous subsection.

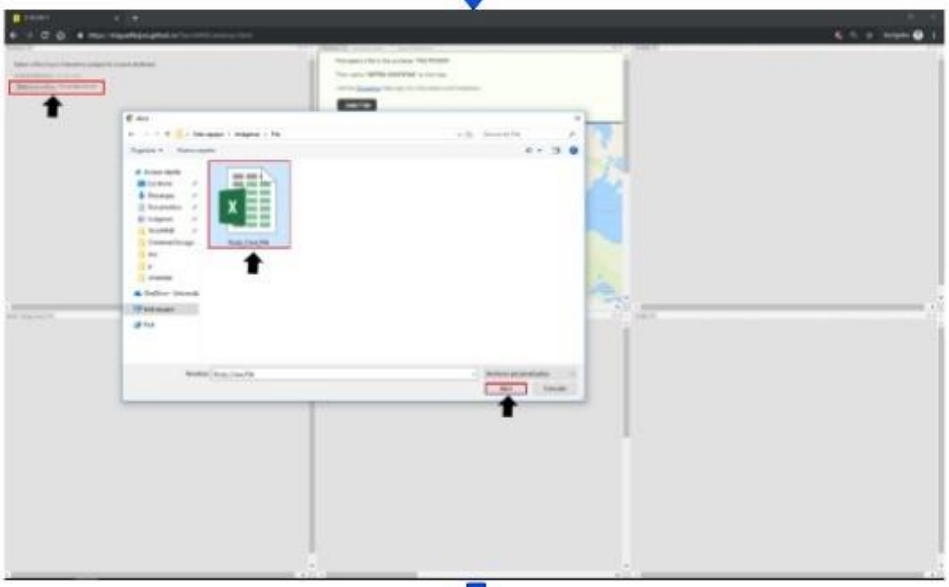
Consider the case in which, from the information provided as a raw file (.xlsx file) by the user, it is desired to know the general proportion of the predominant means of transportation of the people whose reason for traveling is working at night (considered as from 6:00 PM to 6:00 AM), as per genre. Additionally, is desired to analyze in average the stratum, and its neighborhood & locality, and cardinal point in the urban space of the city (North-South-East-West), where these people are concentrated using any means of transportation. On the other hand, is also wanted to know how is its distribution of distance traveled as frequency analysis for all the people living in the same stratum of interest.

Thus, an alternative process to solve the analysis previously described, a user of the tool S-PLOR-T, can develop the with the procedure that the following figure below shows:

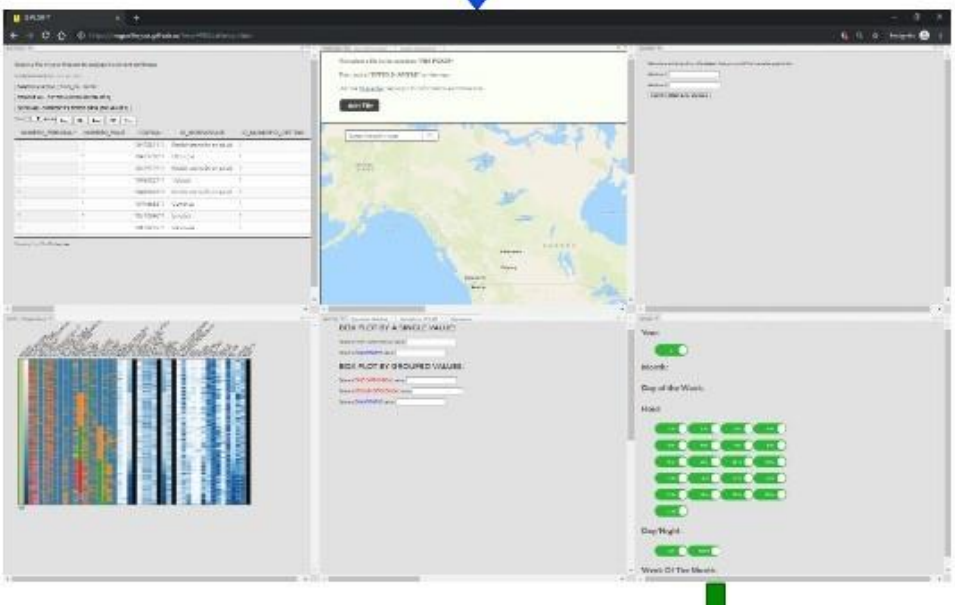
Figure 21 Storyline in S-PLOR-T for the analysis posed as example

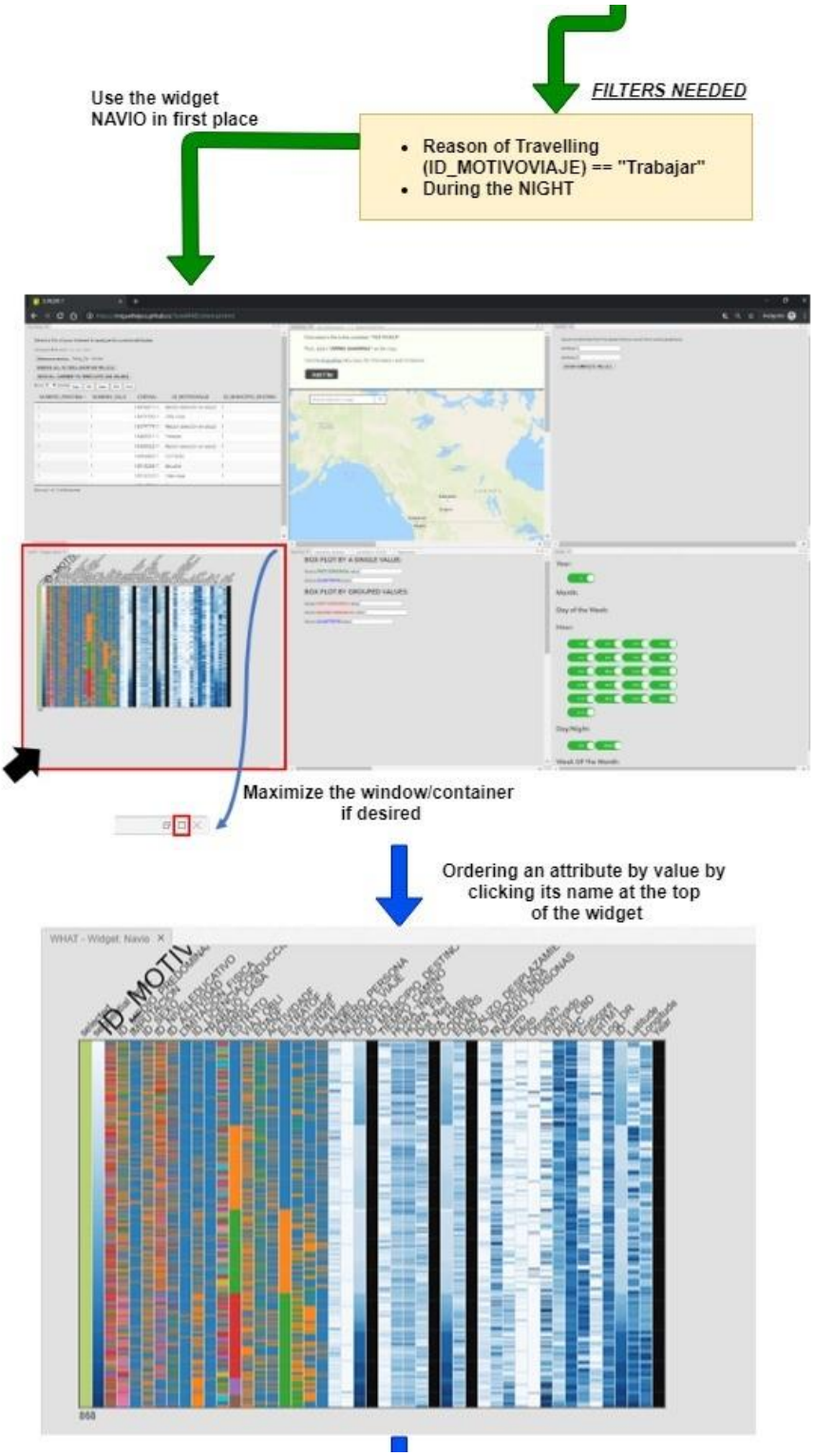


Uploading the file (.xlsx/.csv)

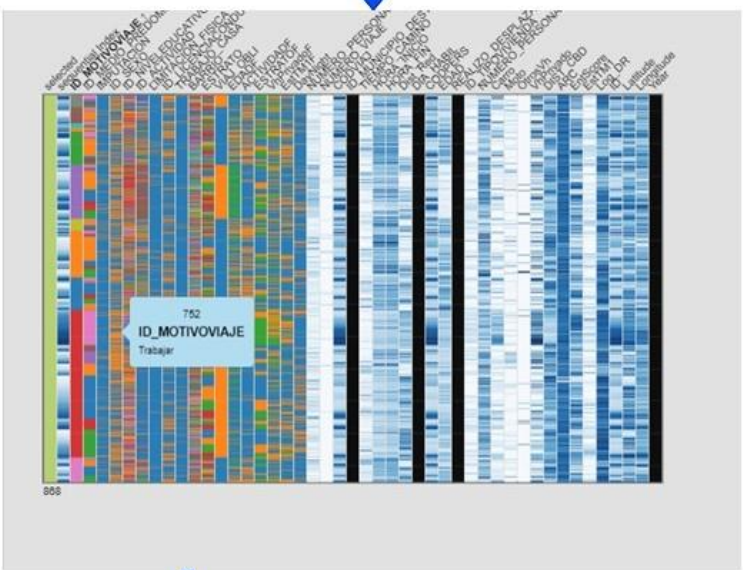


Resulting in ...

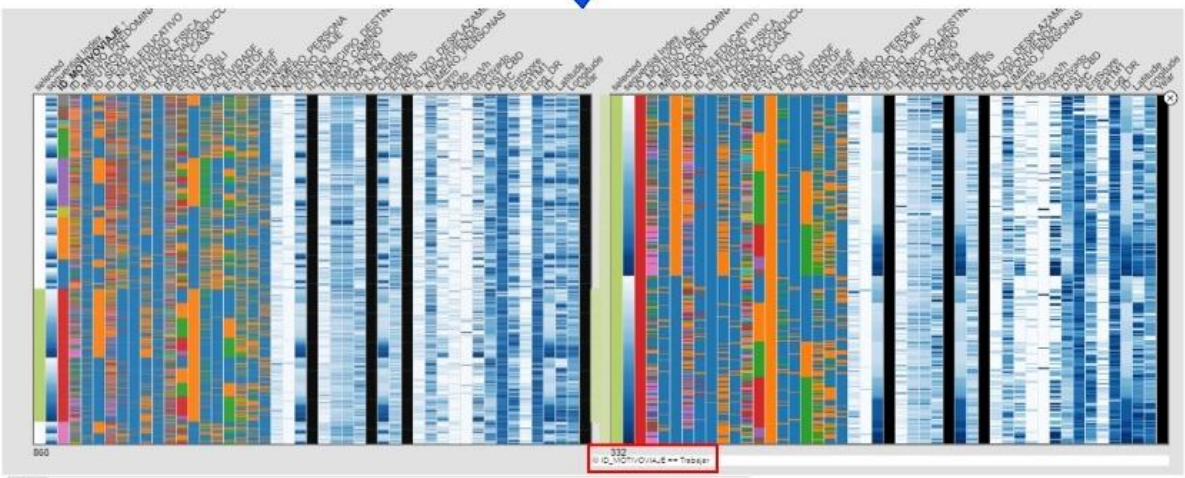


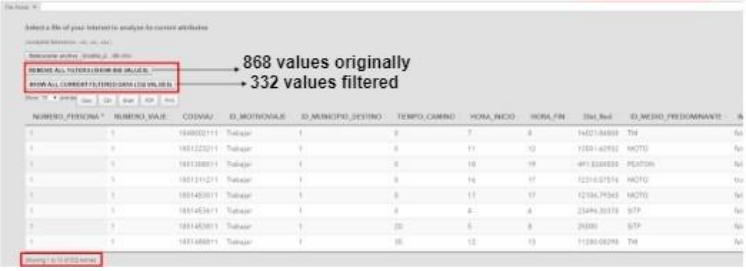

FILTERS NEEDED
Use the widget NAVIO in first place
Reason of Travelling (ID_MOTIVOVIAJE) == "Trabajar"
During the NIGHT
Maximize the window/container if desired
Ordering an attribute by value by clicking its name at the top of the widget
WHAT - Widget: Navio
ID_MOTIV

For filtering the
REASON OF TRAVELLING



Filtering the desired value
by selecting it in the attribute
ordered previously. Will by displayed the filter
as shown in the image.

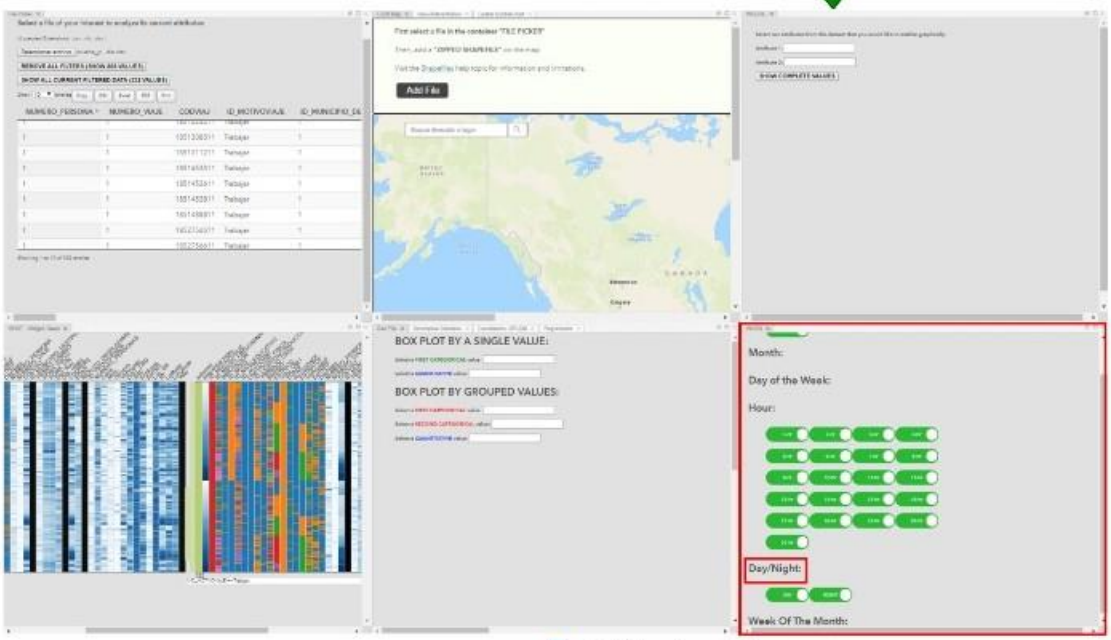




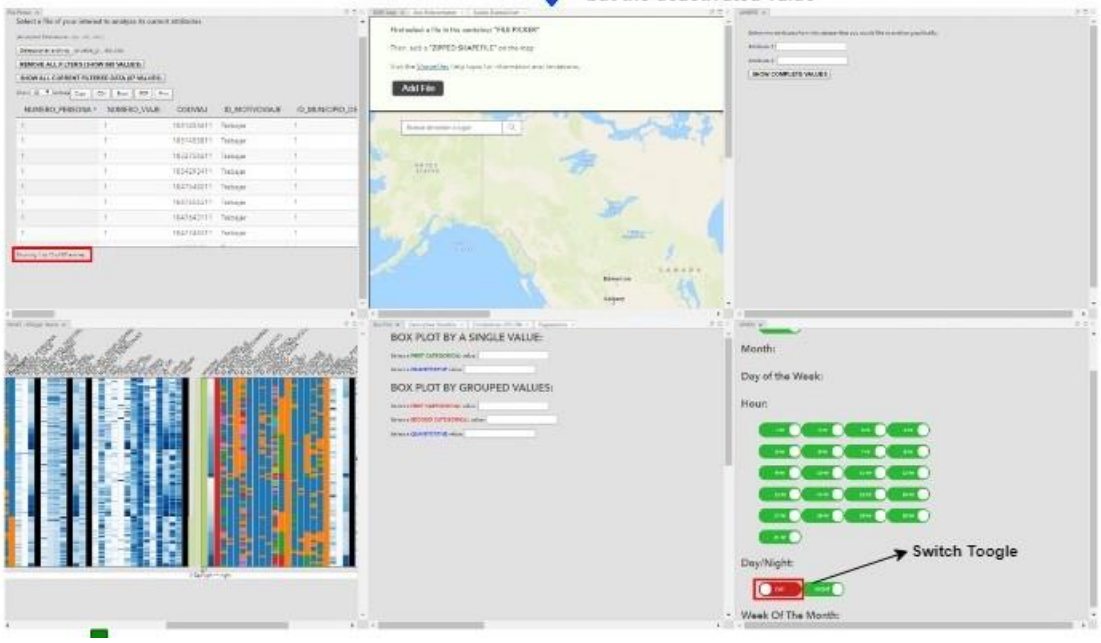
Minimize the container. At Navio's container and the Summary Table's container :
"The data was filtered from 868 values to 332 values (Red Boxes)"

Reading the Preliminary Results

Using the container WHEN can be deactivated the attribute DAY, so that the data only keep a remaining the trips at NIGHT.



Clicking the switch toogle DAY, it will be deactivated. The data will be filtered to all but the deactivated value

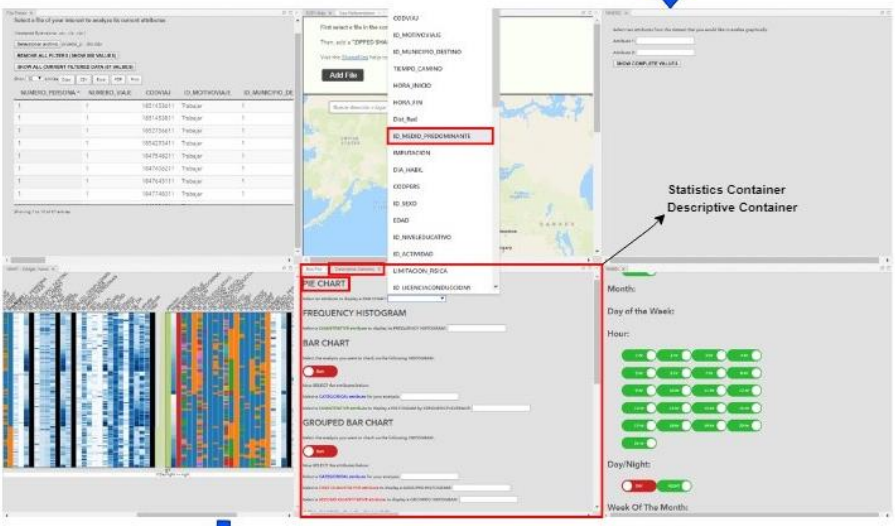


"The data was filtered from 332 values to 87 values (Red Boxes)"

Reading the Preliminary Results

To get the proportion of the predominant means of transportation of the current people:

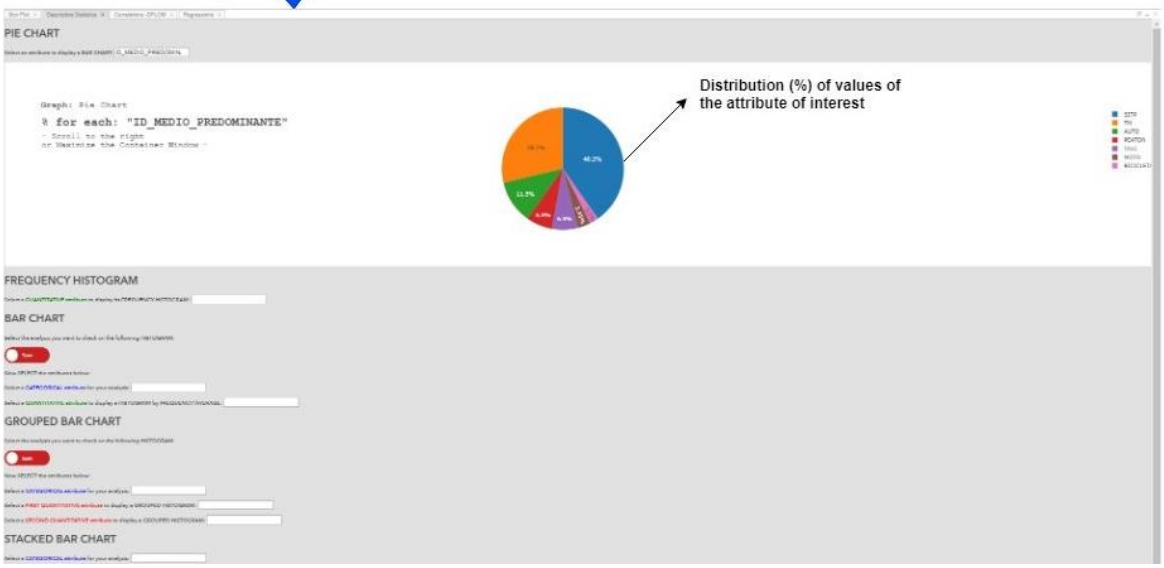
Go to Descriptive Statistics in the Statistics Container and as desiring a PROPOTION (%) use a PIE CHART for this preliminary analysis.



Resulting in ...

```
Graph: Pie Chart
% for each: "ID_MEDIO_PREDOMINANTE"
- Scroll to the right
or Maximize the Container Window -
```

Maximizing the Statistics Container

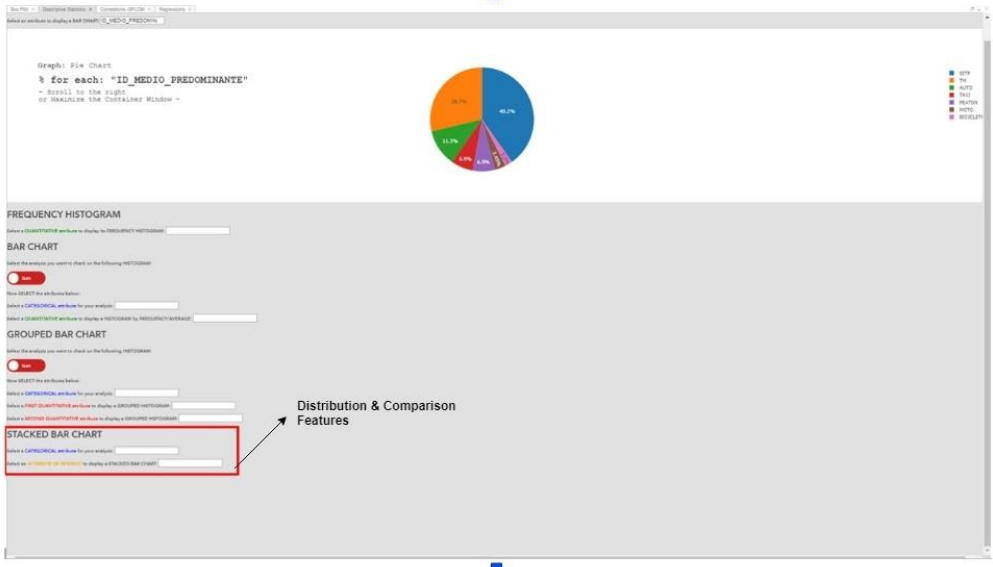


SITP: 35 registers - 40.2%
TM: 25 registers - 28.7%
AUTO: 10 registers - 11.5%
PEATON: 6 registers - 6.9%
TAXI : 6 registers - 6.9%
MOTO: 3 registers - 3.45%
BICICLETA: 2 registers - 2.3%

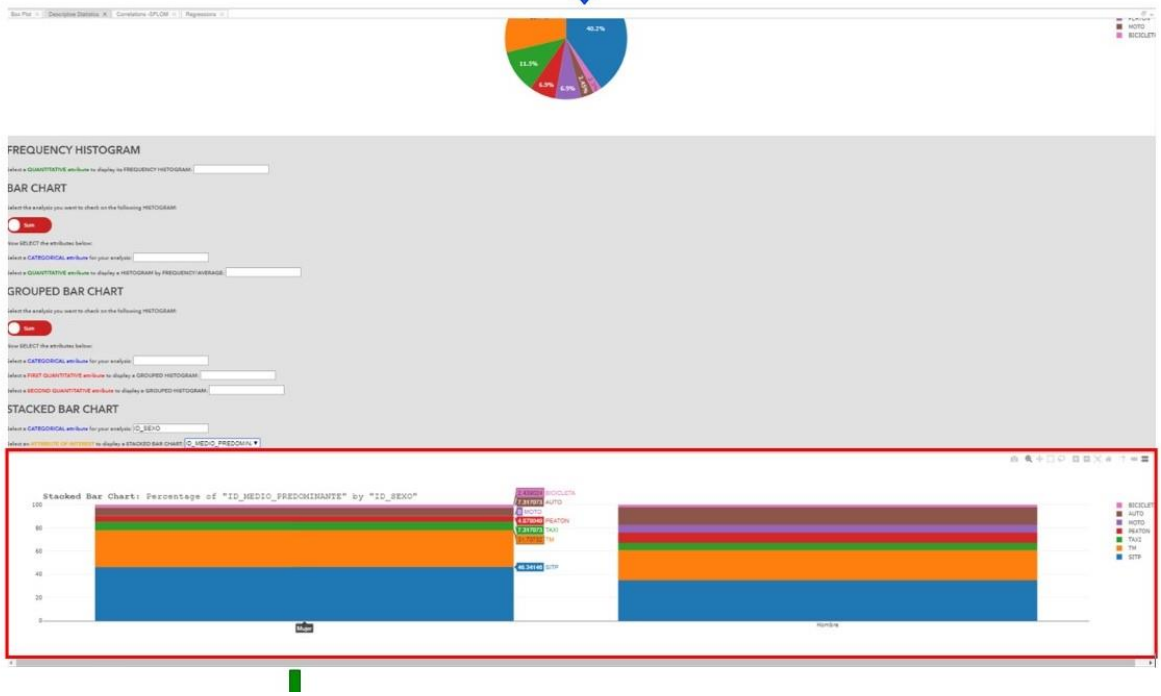
Reading the Preliminary Results

ACTION NEEDED

Now, as needed the information also per genre, the analyst could use a STACKED BAR CHART as displaying information in a percentage scale



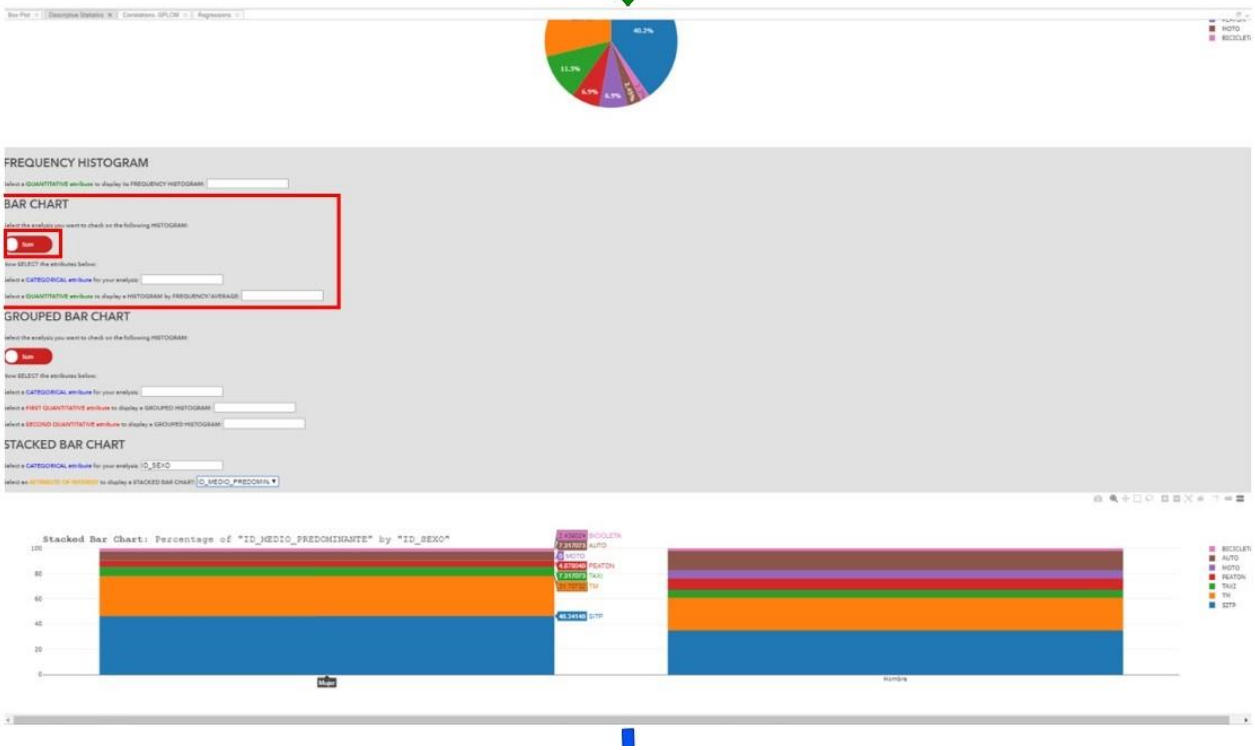
Resulting in ...



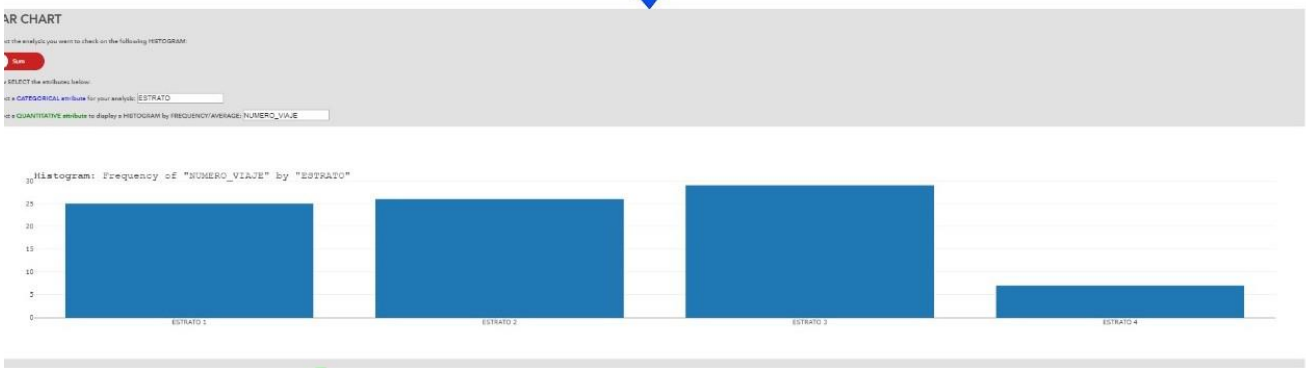
ACTION NEEDED

In order to analyze for the current filters the information to get the stratum that has the biggest frequency of people:

Use a Bar Chart with the toogle switch in SUM.



Selecting the attributes of interest



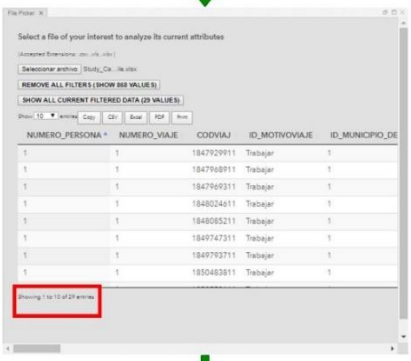
By putting the mouse over any of the bars, will be displayed the detailed information of the frequency calculated per Stratum.

The largest STRATUM is number 3, with 29 registers as frequency.

By clicking on the bar of interest (STRATUM No. 3) the data will be filtered to only those 29 registers.

Reading the Preliminary Results

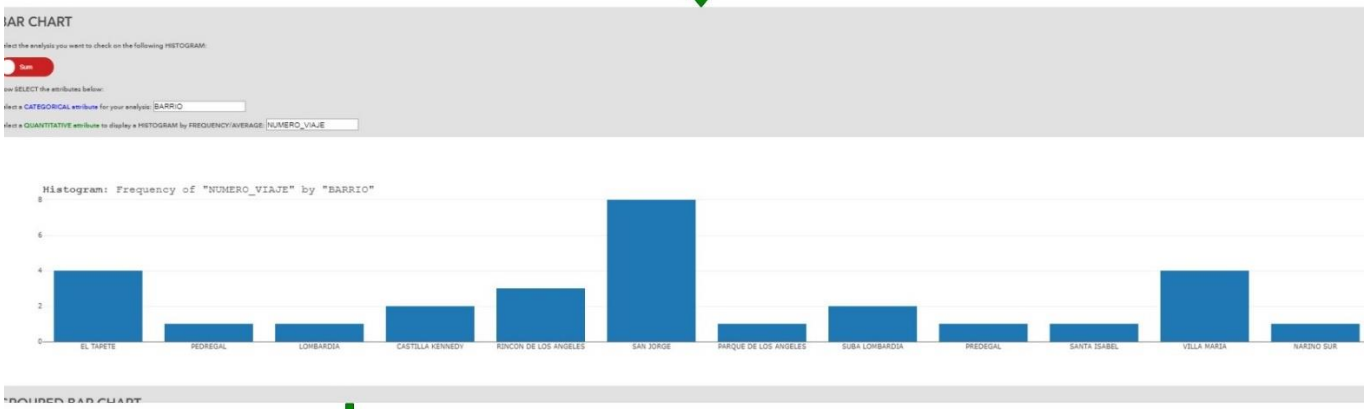
When minimizing the container can be seen the filtered data



Returning to the Container of Statistics

To know the neighborhood which has the biggest frequency of STRATUM 3 and the all current filters, repeat the same process in a BAR CHART BUT SELECTING AS CATHEGORICAL FEATURE the neighborhood

Resulting in ...

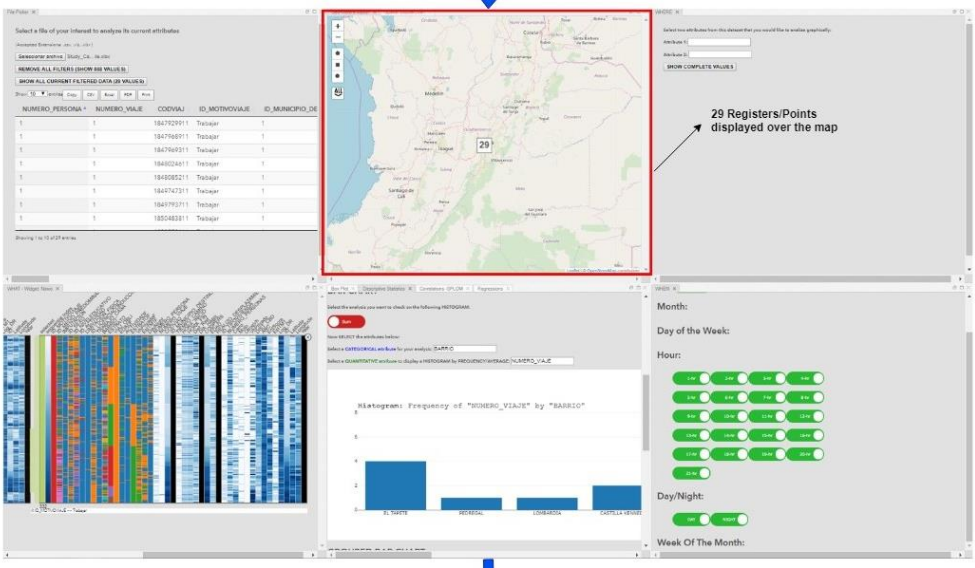


SAN JORGE is the neighborhood that has the biggest frequency of registers with 8 from the 29 recently filtered

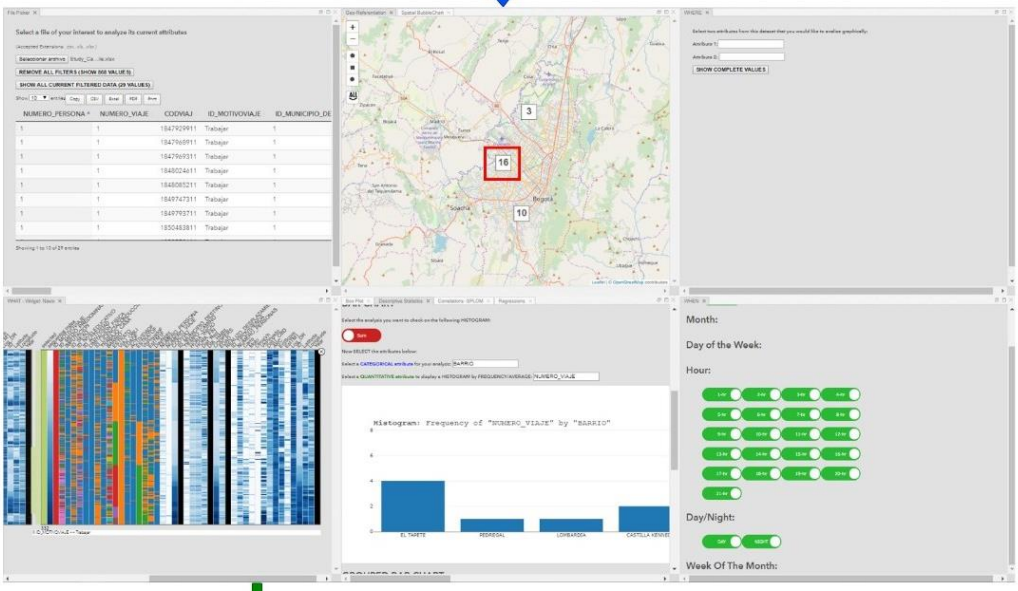
Reading the Preliminary Results

When minimizing the container can be solved the issue of the Cardinal point and the locality which is also desired

Try the SPATIAL ANALYSIS: GeoReferentiation



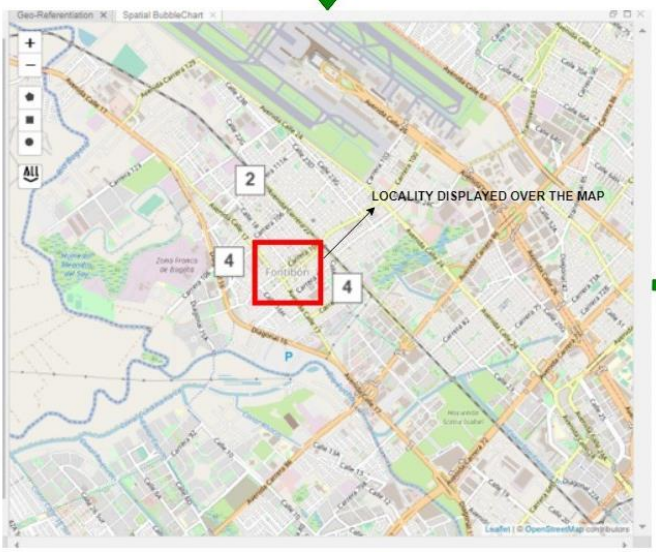
Scrolling over the map will zoom-IN it and will be displayed DETAILED information from the CLUSTERED POINTS



The analyst figures out that all the people that meet with the filters done so far (over the current 29), are concentrated in the cardinal point WEST over the space of Bogotá City.

Reading the Preliminary Results

Zooming-IN more over the map when scrolling ...

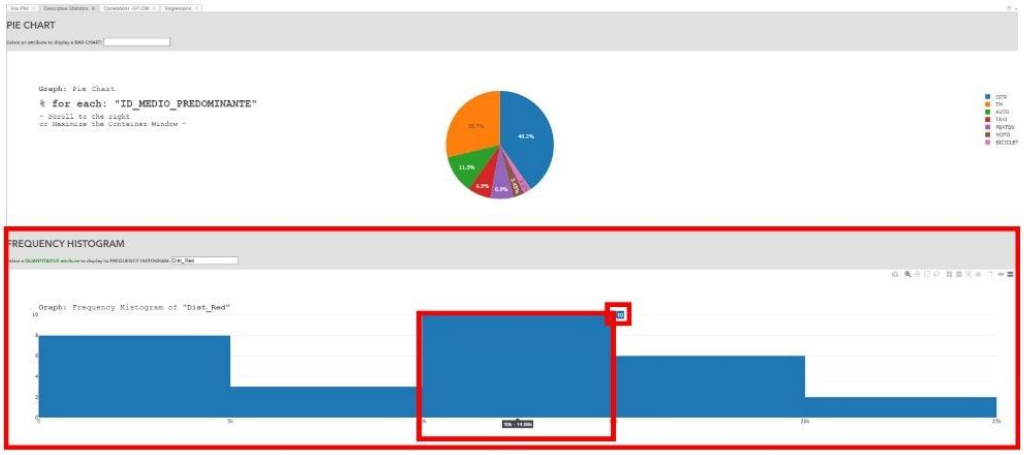


From the map the user can figure out that the Locality where the points are more concnetrated is FONTIBON

As also is wanted to know how is its distribution of distance traveled as frequency analysis for all the people living in the same stratum of interest:

Try a Frequency Analysis by analyzing the DIST_RED for all people living in the STRATUM 3, with all current filters.

Resulting in ....



From which ...

The analyst can figure out that most of the 29 people (10 of them), travels per day almost between 10 and 15 kilometers long. From the analysis can be seen that 62% of the total travels longer than the average, with more than 10 kiometers long.

6. EXPERIMENTAL PHASE

6.1. Used Technologies

For the realization of the present study, the following tools are used:

- **S-PLOR-T v1.0**: Software that supports the exploratory analysis of data with spatial-temporal characteristics, designed and developed within the Project that is summarized through this document. With this tool, the mechanism to evaluate and validate the research question exposed in section 2 of this document is provided. For this, refer to chapter 2: "THE PROBLEM".
- **Microsoft Excel 2016**: This tool is used to provide the participants with the information of the test data to be loaded for the development of the usability evaluation. Additionally, this tool is used for the administration, management and processing of data for the subsequent analysis of these and the proper elaboration of conclusions.
- **Google Chrome V.73.0.3683.103**: Browser by which all the evaluations of this project were done, and which is proved to have its best performance within every analysis provided by it.

6.2. Participating Groups

The description of the groups participating in the project is presented below:

- **Course "*Transportation Systems*" - Universidad de los Andes:**
    - **Program**: Civil Engineering
    - **Level**: Undergraduate
    - **Semester in Curriculum**: 7th Semester
    - **Description**: “The course addresses the principles of traffic engineering and transportation planning. The course provides tools to understand transport in a technical way, within a multidisciplinary framework.”

- **Research Group: "*Group of Studies on Urban and Regional Sustainability - SUR*" - Universidad de los Andes:**
    - **Program**: Civil Engineering
    - **Description**: "Supports the processes of public policy design in the dimensions of urban and territorial ordering. It is characterized by its highly technical and multidisciplinary approach, with lines of research framed in: principles of urban and regional environmental quality; regional competitiveness and logistics; public management and institutionality; urban and regional economy; public space and urban architecture; habitat and sustainable construction; modeling and spatial analysis, and mobility and territorial planning."

6.3. Study Group

The user population that participated in this project was chosen based on the following characteristics:

- **Educational Level**: To be in a level of education of undergraduate or higher. The latter given that it is guaranteed that they possess the skills of analysis and interpretation of data.
- **Age Range**: Be older than 18 years of age.
- **Physiological condition**: The user does not present characteristics related to problems or learning disorders or attention deficit.

The population was obtained through the institutions and academic units that actively participated in the study, as previously described.

6.4. Experimental Methodology

In the first hand, every participant will be introduced to the proposed tool, showing them the generalities of every module/container separately. Through this exercise the participants of the study will be immerse on what the main and general functionalities every component of every container has. This last, in order to contextualize them and letting them know how they could get results from analysis from different and perhaps linked perspectives.

Then, each participant will be presenting an “Usability Evaluation” in an individual manner. This exercise will evaluate, per user, its understanding on the use of the tool to reach the solution of a quantity determined of challenges, that let to interact with various components of the tool as a whole. That is, every challenge permits the interaction with a particular set the attributes of the views proposed to get to the answer of it. Additionally, will be asked about YES/NO questions that are related to the completion by challenge and how it was perceived to be in terms of difficulty. Moreover, comments per challenge are also asked, so that the perception received has more context and description. Time will be set and calculated per challenge per user. All the above, with the purpose of determining the answers got and how were them perceived in general terms.

After the above, once the study is completed, an evaluation of the quality of the tool is made by each of the participants of the test, to indicate the positive and negative aspects that they have experienced, with a view to improving the first version of the tool. Additionally, is also obtained sociodemographic information of each of them, so that can be determined use relationships on the tool S-PLOR-T.

Thus, having the results obtained with the participation of the users participating in the study, we proceed to carry out the due analysis of these, in order to close with the respective conclusions seeking to answer the question posed for the present study.

Below, in the Figure 22, can be appreciated a diagram that outlines the experimental methodology previously described:

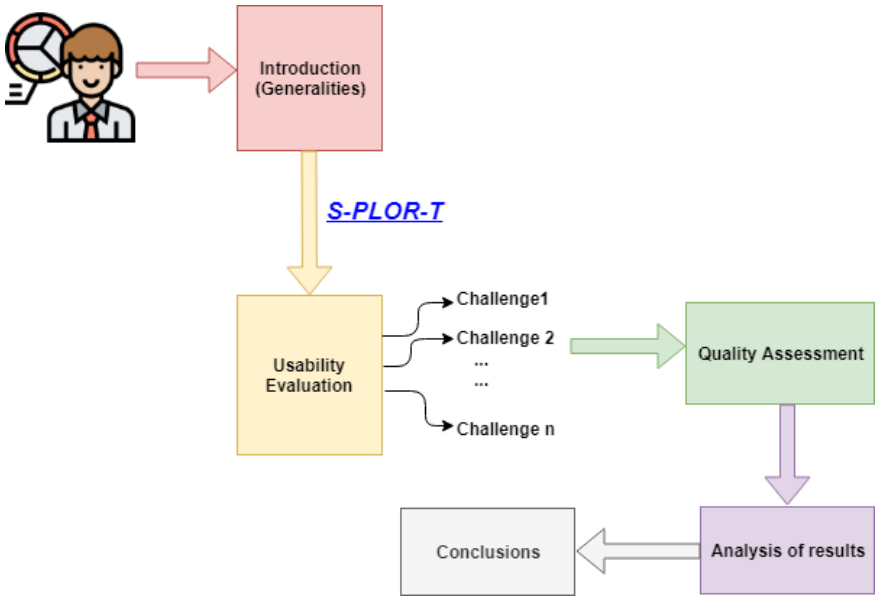


Figure 22 Experimental Methodology

6.5. Evaluation: Study Variables

In order to answer the research question posed previously in this document, a set of variables discriminated by participant user of the proposed tests is defined:

- **Proportion of culminated challenges compared to those proposed by test**
  - **Justification:** This measure will justify the number of challenges considered as completed by all the study participants, compared to the number of challenges initially raised, no matter if the answer given where right or not. This will allow to know how much each participant worked to reach their answers by challenge.
  - **Unit of Measurement**: Percentage
  - **Measuring process:** Every evaluation asks for the completion of each challenge proposed to each participant. That is, no matter what answer the user found through its own analytic process, will be asked

to check a completion YES/NO question, by the perception of each participant on the going. This percentage will be obtained as the average of the percentages, based on the number of answers that were given per challenge by all the participants.

- **Proportion of correct answers registered per Challenge**
  - **Justification:** This measure will justify the number of challenges that were correct by all the study participants in a percentual scope. This will allow to know from how much each participant worked to reach their answers by challenge, how many got the right answer per Challenge.
  - **Unit of Measurement**: Percentage
- **Proportion of the completeness and correctness per Challenge by differencing the participating groups of the study.**
  - **Justification:** This measure will let us determine how the results got during the study differ between the two big participant groups, defined in subsection 6.2. This will allow to know how to define in a better way the use of the tool through how the performance of each group went.
  - **Unit of Measurement**: Percentage

6.6. Quality Assessment for the Tool

This evaluation is presented to those users who worked with the S-PLOR-T tool, once each of the challenges presented in this study was completed. It is done with the purpose that users indicate those positive and negative aspects that they consider important to mention, after having made use of the tool.

For this purpose, they are presented with a group of qualitative questions, with which users are asked to indicate how interesting they seemed to S-PLOR-T, how much they considered that they understood its usability in graphical and analytical terms, and if they consider S-PLOR-T useful for the development of interactive analysis within the paradigm of what, when and where, in a synchronized manner.

Each of the questions presented in this evaluation offers alternatives of options within a LIKERT scale analysis with 6 options, indicating the levels of difficulty and interest of the tool. Additionally, sociodemographic information is asked, which allows to have an idea of the biases in the tests carried out on the different users. In this way, results are provided that can be combined together and properly discriminate the level of acceptance that S-PLOR-T had with the user group that worked with it.

For more details, refer to the Annex in Section 12.3: "Quality Assessment - S-PLOR-T", located in the last section of this document. *It should be noted that since participants were all Spanish-speakers, the perception survey was developed in Spanish.*

6.7. Experimental Phase: Elements and times used

Following the experimental methodology previously discussed in this document and how the usability evaluation of the S-PLOR-T tool was proposed, the times used for the relationship of each stage that was part of this study are indicated below. It should be noted that the times used are independent of the tools used by the students:

- **Introduction (Generalities of the tool)**:
  - **Realization time:** 10-15 minutes
  - Note:
    All general information provided during this exercise is detailed in the user's interaction guide, posed as one annex (Section 12.1).

- **Usability Evaluation**:
  - **Realization Time:** 30-40 minutes (depends on the development of each one of the challenges, per user).
  - **Number of Questions:**
    - For each of the challenges:

- **Challenge No. 1**:
    - 1 Analysis question
    - 2 questions of perception
        - Completeness of the challenge
        - Ease of resolution of the challenge

- **Challenge No. 2**:
    - 1 Analysis question
    - 2 questions of perception
        - Completeness of the challenge
        - Ease of resolution of the challenge

- **Challenge No. 3**:
    - 3 Analysis questions
    - 2 questions of perception
        - Completeness of the challenge
        - Ease of resolution of the challenge

- **Challenge No. 4**:
    - 1 Analysis question
    - 2 questions of perception
        - Completeness of the challenge
        - Ease of resolution of the challenge
- **Challenge No. 5**:
    - 3 Analysis question
    - 2 questions of perception
        - Completeness of the challenge
        - Ease of resolution of the challenge

- **Perception Evaluation and Tool Quality**:
    - **Estimated time:** 5-10 minutes
    - **Number of Questions:**
        - 1 Authorization Question for Collection, tabulation and analysis of results by test
        - 8 Questions with Sociodemographic characteristics (Gender, Age, Role, among others)
        - 11 questions of perception about the tool.

7. PRESENTATION OF THE EXPERIMENTAL RESULTS

Through this subsection, the results obtained from the usability evaluation and the survey about the quality assessment of the tool carried out with the participating users are presented. These is the result obtained with each participant of the evaluation following the experimental methodology designed and described in the subsection above. For checking out the evaluation presented by the users, refer to Section 12.2: "*Usability Evaluation*".

7.1. Experimental / Technical Evaluation: Obtained Results

The results presented below are based on the variables defined for the validation of this study.

As can be seen in the following table and graph, after carrying out this study it is observed that the participants opted to define as complete the challenge posed in its majority. That is, it was defined as an answer for completeness: "YES". Similarly, can also be seen a pattern of increase in the positive response from the participants, possibly for reasons of ease of challenge or understanding of the tool as going deeper into it.

COMPLETENESS	# Participants	47			
	Challenge	SI	NO	NaN	Total
	Ch_1	78,72%	19,15%	2,13%	100,00%
	Ch_2	95,74%	4,26%	0,00%	100,00%
	Ch_3	82,98%	6,38%	10,64%	100,00%
	Ch_4	57,45%	14,89%	27,66%	100,00%
	Ch_5	23,40%	0,00%	76,60%	100,00%

Table 2 Percentual Summary: Completeness

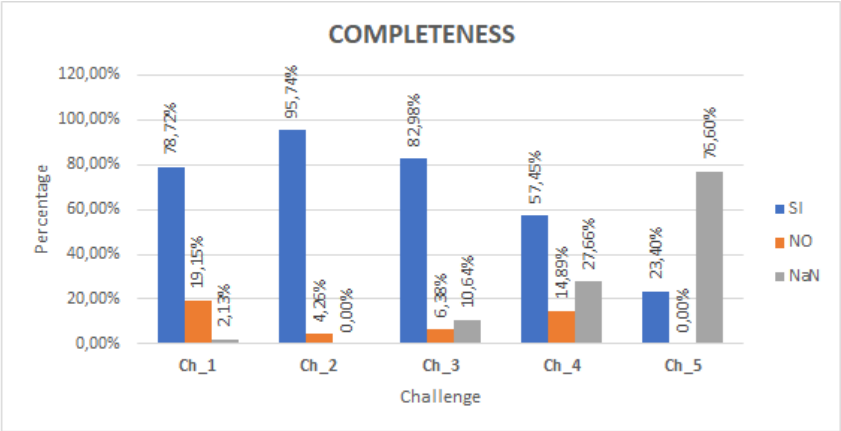


Figure 23 Distribution of the Completeness by Challenge

In this order of ideas, we can determine a general pattern of how the behavior of the participants was on determining if the challenges were considered as "complete". In general terms about the completeness, a 67.658% in average of the whole 47 evaluated answered YES (SI), an 8.936% answered NO (NO) and a 23.406% did not answer the question.

On the other hand, as can be seen in the following table and graph, after carrying out this study it is observed that the participants with a 53.616% in average considered that was easy to solve the challenges posed in the evaluation. That is, it was defined as an answer for difficulty: "YES".

DIFFICULTY	# Participants	47			
	Challenge	SI	NO	NaN	Total
	Ch_1	57,45%	23,40%	19,15%	100,00%
	Ch_2	89,36%	6,38%	4,26%	100,00%
	Ch_3	59,57%	23,40%	17,02%	100,00%
	Ch_4	44,68%	12,77%	42,55%	100,00%
	Ch_5	17,02%	6,38%	76,60%	100,00%

Table 3 Percentual Summary: Difficulty

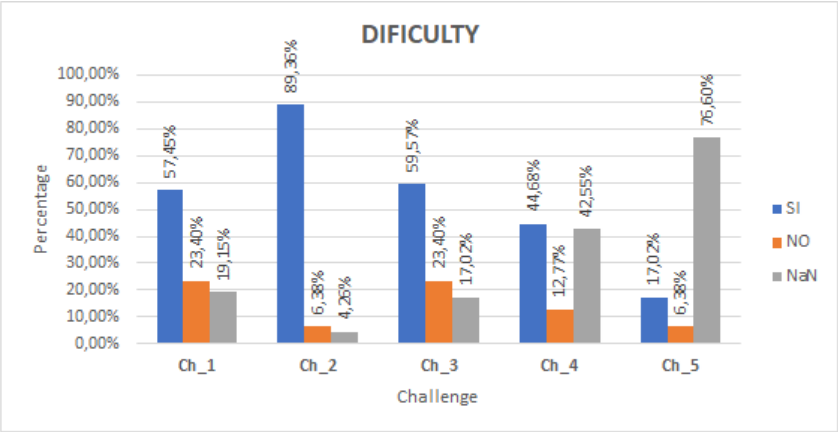


Figure 24 Distribution of the Difficulty by Challenge

Additionally, following the analysis by Challenge posed, can be seen below the standard deviations from the answers provided by all the participants (47):

STANDARD DEVIATION		Completeness
		Standard Deviation
	Ch_1	18.90326251
	Ch_2	25.42308662
	Ch_3	20.23198787
	Ch_4	10.26320288
	Ch_5	18.44812547

STANDARD DEVIATION		Difficulty
		Standard Deviation
	Ch_1	2.872036782
	Ch_2	1.410681977
	Ch_3	1.956835017
	Ch_4	1.53225407
	Ch_5	1.119704146

Table 4 Standard Deviations for study variables: Completeness & Difficulty

From the standard deviations presented above, for both study variables, is clear how the distribution does not follow a Normal curve, which let us affirm that through all the answers is represented a variation. This, in a comparative manner between both variables, shows how between the participants is varied the answers in a bigger way in terms of completeness more than the difficulty. That is possibly due to the perception or perhaps the level of understanding that each participant has as an individual when answering each challenge posed.

However, notwithstanding the last result's description in terms of completeness and difficulty, it is important to clarify that between the participant groups evaluated can be appreciated a huge difference that lies in the educational level or level of expertise of each. In a comparative way, can be seen this difference when analyzing both participating groups in an individual manner, as follows:

COMPLETENESS (EXPERTS - GRADUATES "RESEARCH GROUP *SUR*")	# Participants	11			
	Challenge	SI	NO	NaN	Total
	Ch_1	100.00%	0.00%	0.00%	100.00%
	Ch_2	100.00%	0.00%	0.00%	100.00%
	Ch_3	100.00%	0.00%	0.00%	100.00%
	Ch_4	100.00%	0.00%	0.00%	100.00%
	Ch_5	100.00%	0.00%	0.00%	100.00%

COMPLETENESS (BACHELOR STUDENTS)	# Participants	36			
	Challenge	SI	NO	NaN	Total
	Ch_1	72.22%	25.00%	2.78%	100.00%
	Ch_2	94.44%	5.56%	0.00%	100.00%
	Ch_3	77.78%	8.33%	13.89%	100.00%
	Ch_4	44.44%	19.44%	36.11%	100.00%
	Ch_5	0.00%	0.00%	100.00%	100.00%

Table 5 Comparative Percentual Tables for study variable: Completeness – for Experts and Undergraduate Students separately

DIFFICULTY (EXPERTS - GRADUATES "RESEARCH GROUP SUR")	# Participants	11			
	Challenge	SI	NO	NaN	Total
	Ch_1	100.00%	0.00%	0.00%	100.00%
	Ch_2	90.91%	9.09%	0.00%	100.00%
	Ch_3	100.00%	0.00%	0.00%	100.00%
	Ch_4	100.00%	0.00%	0.00%	100.00%
	Ch_5	72.73%	27.27%	0.00%	100.00%

DIFFICULTY (BACHELOR STUDENTS)	# Participants	36			
	Challenge	SI	NO	NaN	Total
	Ch_1	44.44%	30.56%	25.00%	100.00%
	Ch_2	88.89%	5.56%	5.56%	100.00%
	Ch_3	47.22%	30.56%	22.22%	100.00%
	Ch_4	27.78%	16.67%	55.56%	100.00%
	Ch_5	0.00%	0.00%	100.00%	100.00%

Table 6 Comparative Percentual Tables for study variable: Difficulty – for Experts and Undergraduate Students separately

As presented in the comparative tables above, in terms of both study variables, is clear how the appreciations given by both participating groups differ. That could be possibly due to the expertise that the participant has, in terms not only in an analytical manner but in knowledge such as on the kind of data used in the user test. Moreover, considering only the experts, who are the members of the research group SUR from Universidad de los Andes, the considerations about these same terms of completeness and difficulty seems that they were very receptive in each challenge, understanding and executing what was questioned in a positive manner.

On the other hand, as follows in the table and graph below, another analysis of great interest apart from the completeness and perception in terms of difficulty by challenge, is the one following the proportion of correctness for each in comparison with the whole posed. From the following, can be appreciated that all the participants had a 49.504% in average for correct answers through all the challenges posed. However, even though it seems not to be a good result about correctness in average for all the challenges posed, if considering just the first four challenges (as the undergraduate students did not

considered the solving of the fifth challenged posed), a 56.915% in average for correct answers through all of them is got.

CORRECTNESS	# Participants	47		
	Challenge	CORRECT	INCORRECT	Total
	Ch_1	57.45%	42.55%	100.00%
	Ch_2	74.47%	25.53%	100.00%
	Ch_3	44.68%	55.32%	100.00%
	Ch_4	51.06%	48.94%	100.00%
	Ch_5	19.86%	80.14%	100.00%

Table 7 Percentual Summary: Correctness

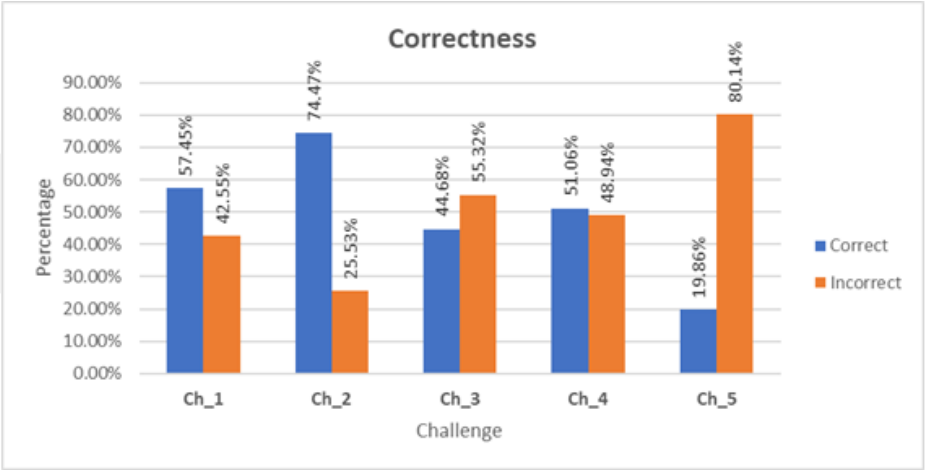


Figure 25 Distribution of the Correctness by Challenge

The above, can be perhaps in terms of the difficulty of each challenge, or the understanding of the terms and visualizations and interactions needed per challenge by participant.

Now, following the same order of ideas for the two previous study variables, a comparison between the two major participating groups can be observed below:

CORRECTNESS (BACHELOR STUDENTS)	# Participants	36		
	Challenge	CORRECT	INCORRECT	Total
	Ch_1	44.44%	55.56%	100.00%
	Ch_2	66.67%	33.33%	100.00%
	Ch_3	32.41%	67.59%	100.00%
	Ch_4	36.11%	63.89%	100.00%
	Ch_5	0.00%	100.00%	100.00%

CORRECTNESS (EXPERTS - GRADUATES "RESEARCH GROUP SUR")	# Participants	11		
	Challenge	CORRECT	INCORRECT	Total
	Ch_1	100.00%	0.00%	100.00%
	Ch_2	100.00%	0.00%	100.00%
	Ch_3	84.85%	15.15%	100.00%
	Ch_4	100.00%	0.00%	100.00%
	Ch_5	84.85%	15.15%	100.00%

Table 8 Comparative Percentual Tables for study variable: Correctness – for Experts and Undergraduate Students separately

From the above, you can see the same behavior as in the previous comparison, so that the experts had a better performance in responding to each challenge than the undergraduate students. Perhaps it is due to the interest of the participants in solving the user test or due to the understanding of how the tool works to obtain a correct result and analysis when interacting between the variants of the modules presented in the first version of the application (prototype). However, it can also be seen that for most of the challenges, considering the resolution of the challenges that all the participants fully developed (from challenge 1 to 4), analyzing results of correctness by challenge, the

correct results exceed the population average considered for the present study. The above is a positive implication for the tool, being evident that the participating users (experts and non-experts) in the course of the test, became familiar with the tool, understanding the location of the graphic elements and how to carry out the various analyzes proposed.

Furthermore, in the following representations of the time invested by challenge, the variations in data taken, per participant per challenge can be appreciated, given the differentiation between the study groups: experts versus the entire population studied. This, as an affirmation through how the standard deviations differ for both cases, being lower for the expert participants in the present study.

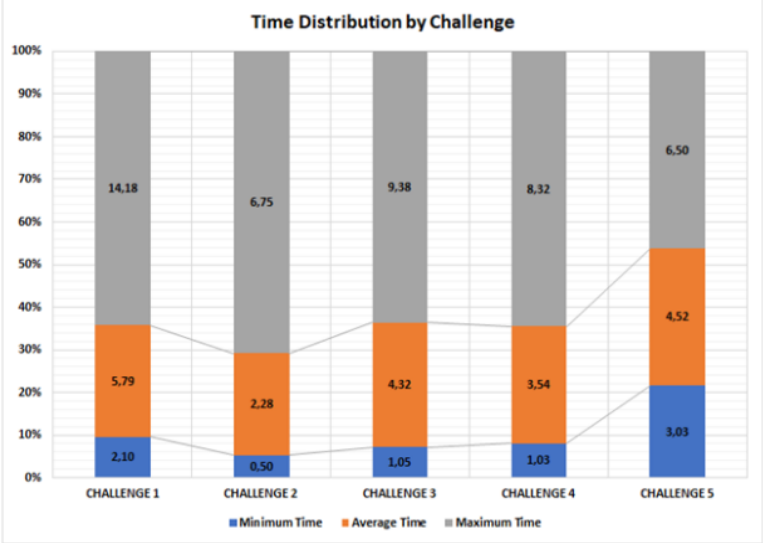
		Time Invested (min)	
	Challenge	Average	Standard Deviation
STANDARD DEVIATION	Ch_1	5.79	2.872036782
	Ch_2	2.28	1.410681977
	Ch_3	4.32	1.956835017
	Ch_4	3.54	1.53225407
	Ch_5	4.52	1.119704146

		Time Invested (min)	
	Challenge	Average	Standard Deviation
STANDARD DEVIATION (EXPERTS-GRADUATES)	Ch_1	3.66	1.377966955
	Ch_2	1.83	0.906931779
	Ch_3	3.74	1.211779308
	Ch_4	2.59	0.872552855
	Ch_5	4.52	1.119704146

Table 9 Comparative Standard Deviation for Time Invested by Challenge – for Experts and Undergraduate Students separately

That is, as evidenced in the following graphs, the times invested differ between the two participating groups in the tests. The last can be appreciated, by looking at the behavior of the variation of average times between challenges, particularly for users with a certain quality of expertise. However, in general terms, there is also evidence of understanding and familiarization of the tool, as the times per challenge decrease (in different proportions between groups). Even if this pattern is evident, an analysis can also be made that implies that the tool was much better received in terms of the time required by analysis by all the expert users, who obtained the best results in time.

Figure 28 Time Distribution by Challenge



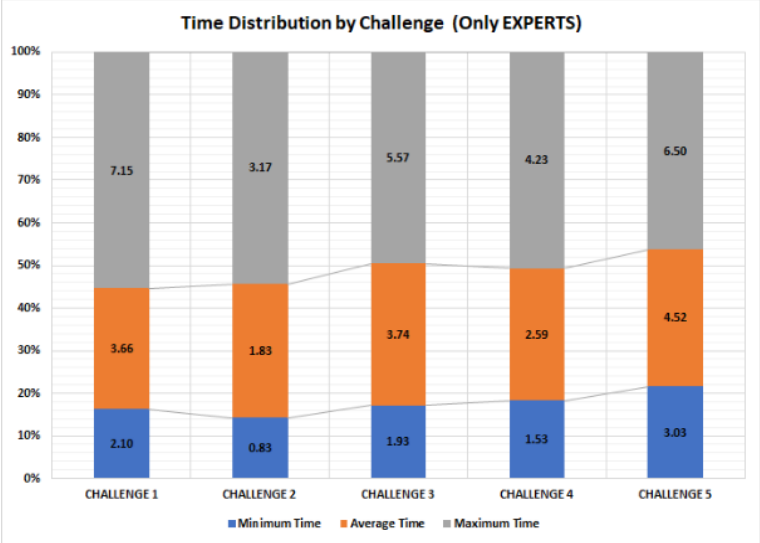


Figure 27 Time Distribution by Challenge for only Experts

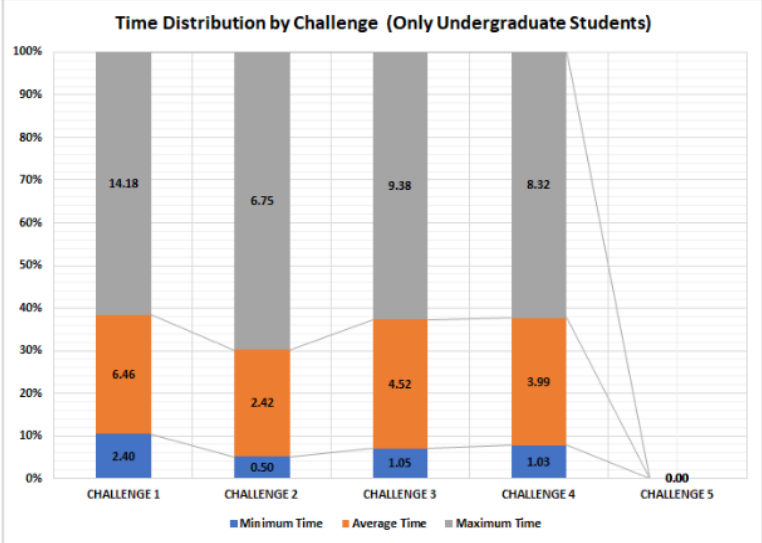


Figure 26 Time Distribution by Challenge for only Undergraduate students

If viewed from a different graphical perspective, considering the average times for the entire population participating in the study, familiarization with the tool previously mentioned in terms of time is evident. That is to say, it can be observed that for a majority of individual user evaluations, as the challenges are happening, the time invested in each of them decreases. In addition, it is also evident the amount of challenges that alter the results of analyzed times, considering that users could not respond to all the challenges. This is evidenced in the graph that is briefly presented below:

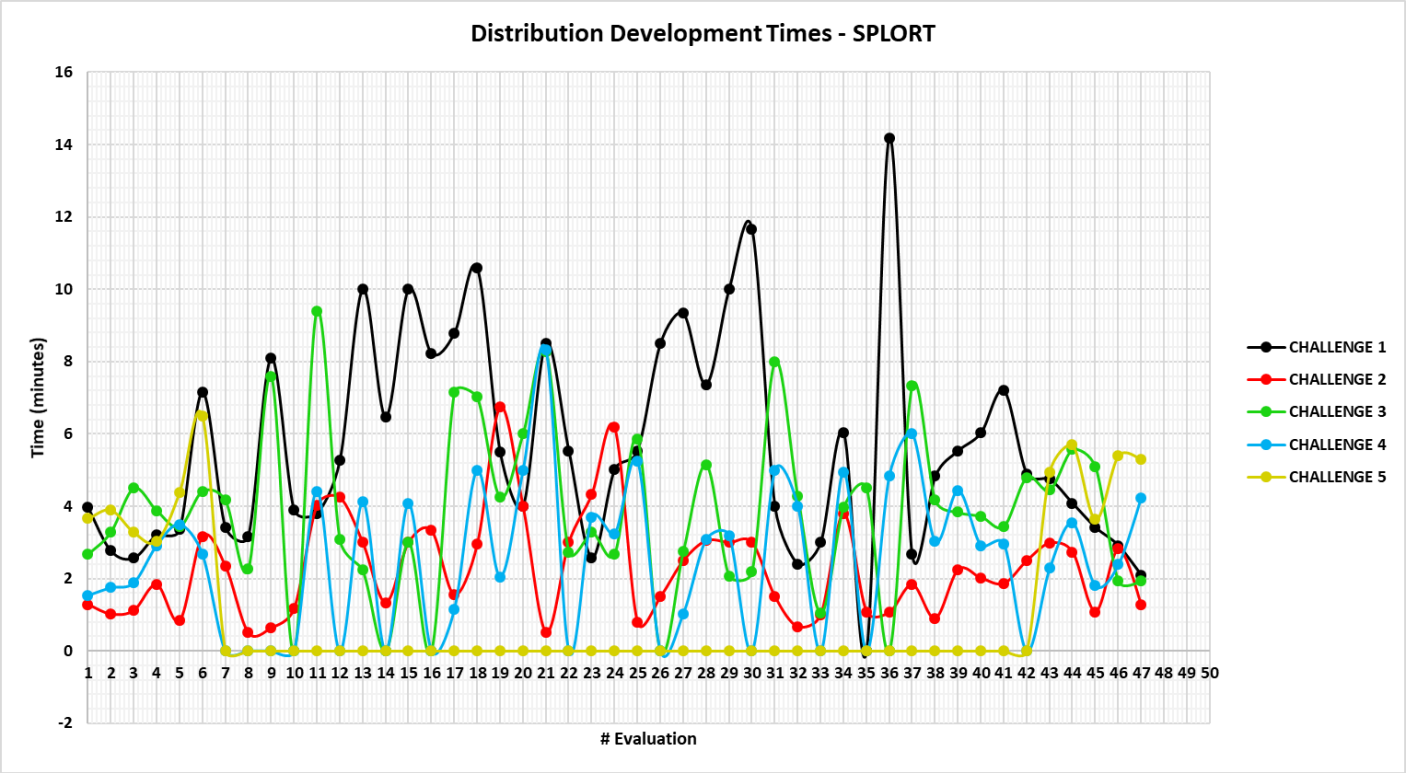


Figure 29 Comparative Plot for Invested Times per Challenge per Participant

7.2. Quality Assessment: Obtained Results

Below are presented the results obtained with the qualitative questions of perception about the tool that the participating users of the study answered once the usability test designed and proposed, for research purposes, was completed. Based on these results, it is expected to understand the experience of all the participating users when interacting with the tool for the purposes of exploratory data analytics.

As it has been evident in the review of the general results of the tests carried out on the participants, and as the following set of graphs shows, in demographic terms there is a greater number of undergraduate students, belonging to the subject of Transportation Systems of Civil Engineering, where one of the tests was carried out (36 participants). The remaining population are the experts with some level of expertise in analytical issues and in terms of knowledge of the data (11 participants).

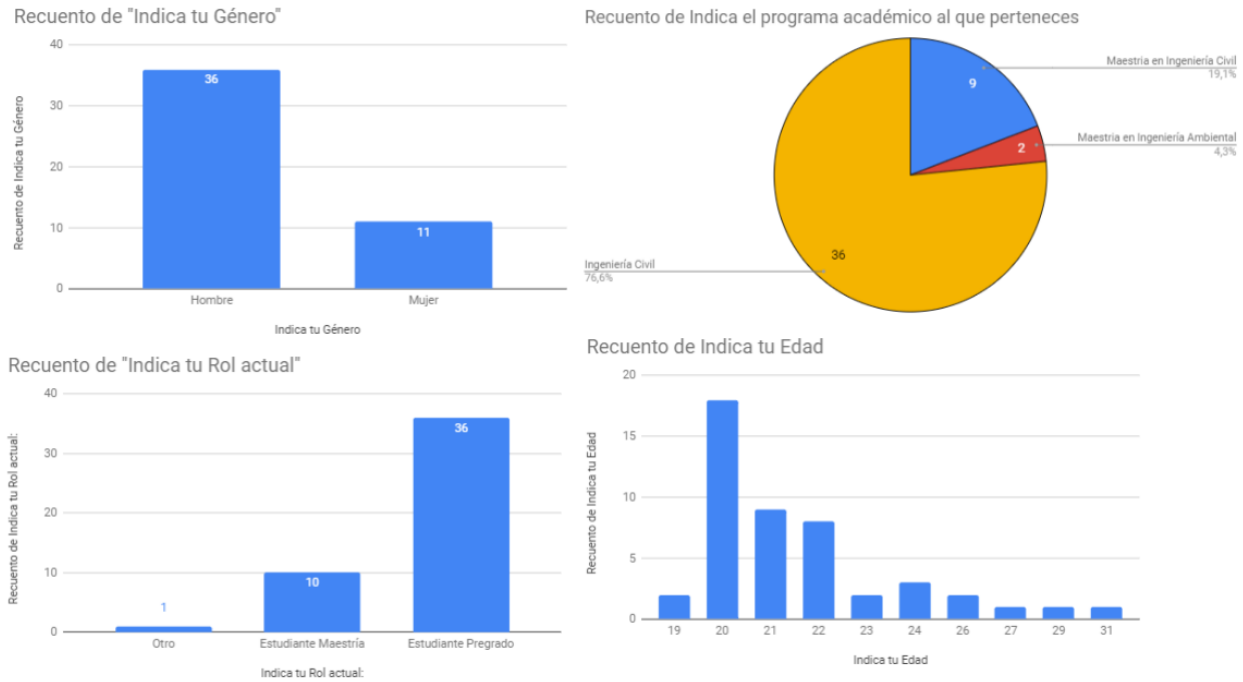


Figure 30 Distribution of the Demographic Information

Now, to have knowledge about the context in terms of data analytics in general terms and, in particular, about the knowledge of the study participants in particular in their consideration of the domain of spatiotemporal data, we also chose to ask each user about their experience and domain knowledge. The results on this concern are shown in the set of graphs presented below.

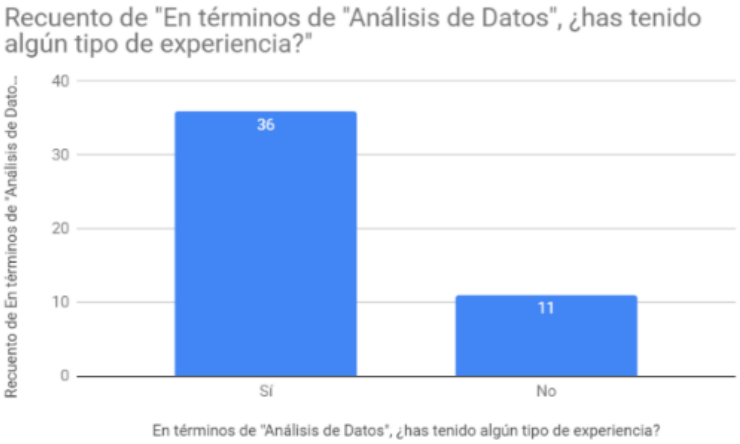


Figure 32 Distribution for: "Experience in Data Analytics"

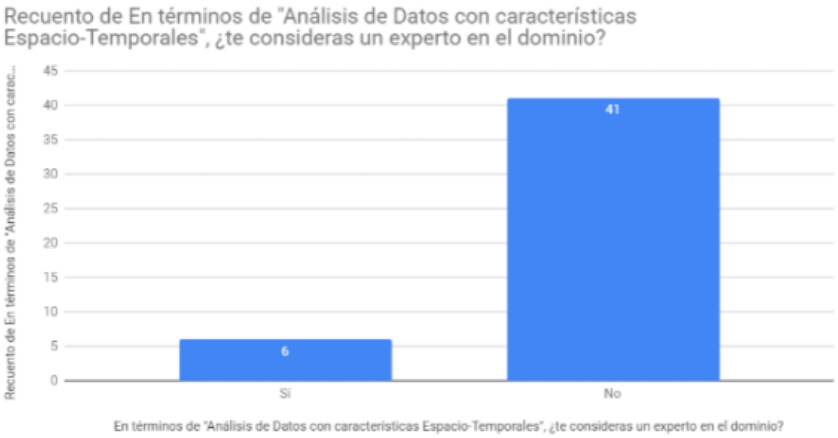


Figure 31 Distribution for: "Expertise in Data Analytics for Spatiotemporal Data"

From the above, it can be stated that of the entire population participating in the study, 76.59% (equivalent to 36 people) has had some experience in data analysis but does not

consider having expertise in the domain of data analytics with spatiotemporal characteristics (87.23%). Well, if the question of expertise can become general for a correct interpretation of said brief result, it aims to link it with the results of resolution of the challenges posed for this study, as presented in the previous subsection.

Now, in specific terms, perceptual questions were asked, as shown below, where results are sought on a scale of 0 to 5 to perform a Likert analysis as shown later in this section.

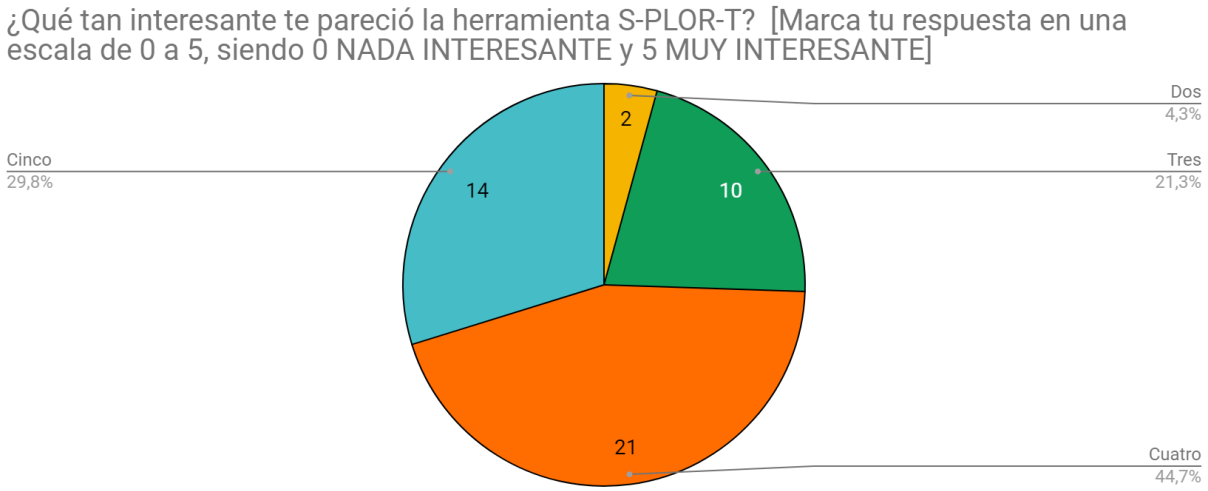


Figure 33 Percentual Distribution for: "Interesting for participants"

In this order of ideas, as can be seen in the previous pie chart, it was intended to find patterns of perception in terms of the interest of the participating users about the tool they had just used. From the brief results got from this previous chart, is evident that considering in the 6-position scale used for this question the answers Zero (0) to Two (2) as negative and Three (3) to Five (5) as positive, in general terms the participants perceived S-PLOR-T interesting in a 95.74% (45 participants), which let us have an idea about the motivation they could have had when using the web tool posed.

Figure 35 Percentual Distribution for:" Ease of use in graphics terms"

Figure 34 Percentual Distribution for:" Ease of use in analytic terms"

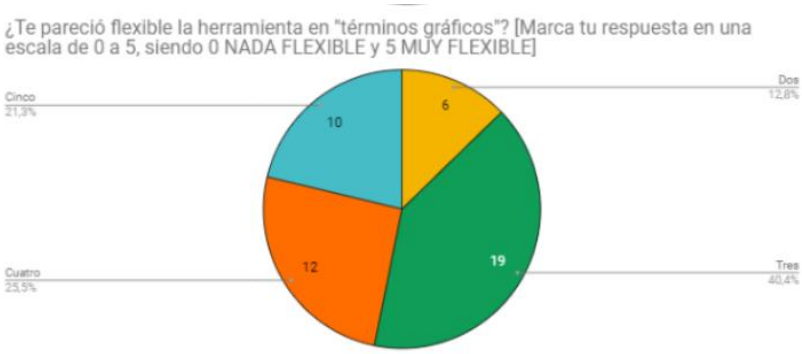


Figure 39 Percentual Distribution for:" Flexibility of use in graphics terms"

Figure 38 Percentual Distribution for:" Flexibility of use in graphics terms"

Figure 37 Percentual Distribution for:" Ease of content's location and functionalities"

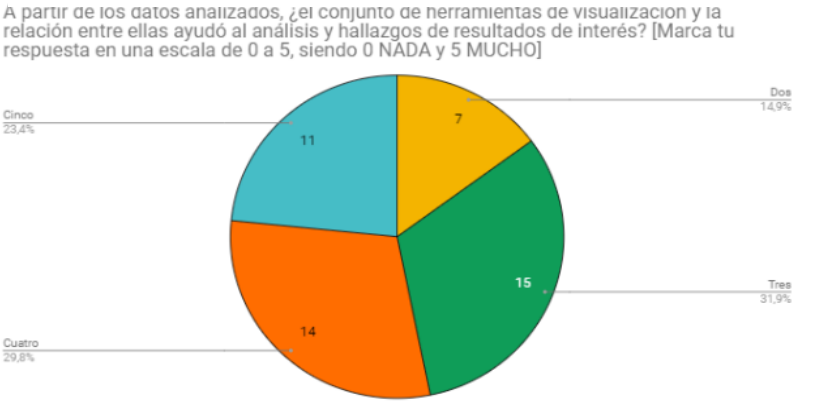


Figure 36 Percentual Distribution for:" General Usability Perception of the tool"

As shown in the set of previous pie charts, in general terms, several aspects were asked in terms of graphics and analytical terms. Analyzing the results in pairs of similar questions, it is consistent that approximately the same number of people participating in the study find the tool used either easy or flexible. In terms of graphic ease of use, 70.21% provided a positive response (between 3 and 5). Similarly, 72.34% approve the tool in terms of ease of use when talking about data analytics. However, it must be taken into account that 29.78% for ease in graphic terms and 27.6% in analytical terms, considered to respond as "slightly positive", which gives rise to see opportunities for improvement to the tool to provide the user with an intuitive tool that allows easy use from different perspectives.

On the other hand, talking in terms of flexibility in graphic terms, 87.23% considered responding positively. In this way, following the same idea as in the previous analysis, 40.4% considered the tool "slightly flexible", perhaps following the same logic of ease of use in graphic terms, which is consistent, and 25.53% in terms of analytics. of data. The last result is a positive result, being an indication that for the population participating in the study, S-PLOR-T is a tool that can provide data analysis from different perspectives to the user's needs, being a 72.34% the answers ranked as positive in the 6-position scale.

Notwithstanding all the previous analysis, it is important to emphasize that the link between ease of use and flexibility of the tool must be fully complied, being these brief results crucial to find opportunities to improve it until its best performance. This is justified by the analysis taken from the perception question if it was easy to locate the content and functionalities of the tool, where 68% answered positively. However, the difference considered it difficult to manage. Furthermore, despite the last fact, 85.10% considered that the set of visualization tools and the relationship between them helped the analysis and findings of results of interest.

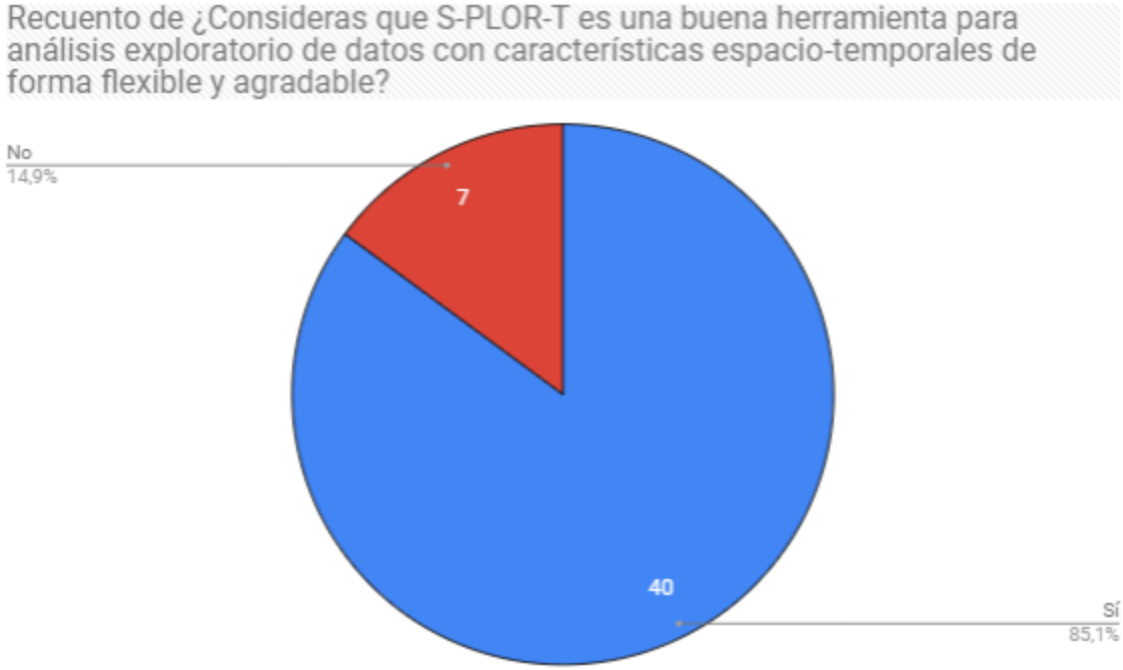


Figure 40 Percentual Distribution for:"Perception of S-PLOR-T as an exploratory tool for data analytics"

On the other hand, it is a very good result that of the totality of the participating population, 85.1% consider S-PLOR-T as a good tool for exploratory data analysis with spatiotemporal characteristics in a flexible and pleasant way. The above, allows to see that the tool was well received by users and that for the purposes for which the tool is designed, expectations are met.

It should be clarified that as variation was shown in the results between the two groups participating in the study in the usability evaluation (user test), in the perception results there is also significant variation, which allows to determine and define in a better way the target population of this tool (e.g. by level of expertise). Below can be appreciated a Likert plot which let to analyze this pattern between both participating groups through all the "Quality Assessment of the tool" survey.

- **Likert Analysis over 6 positions:**

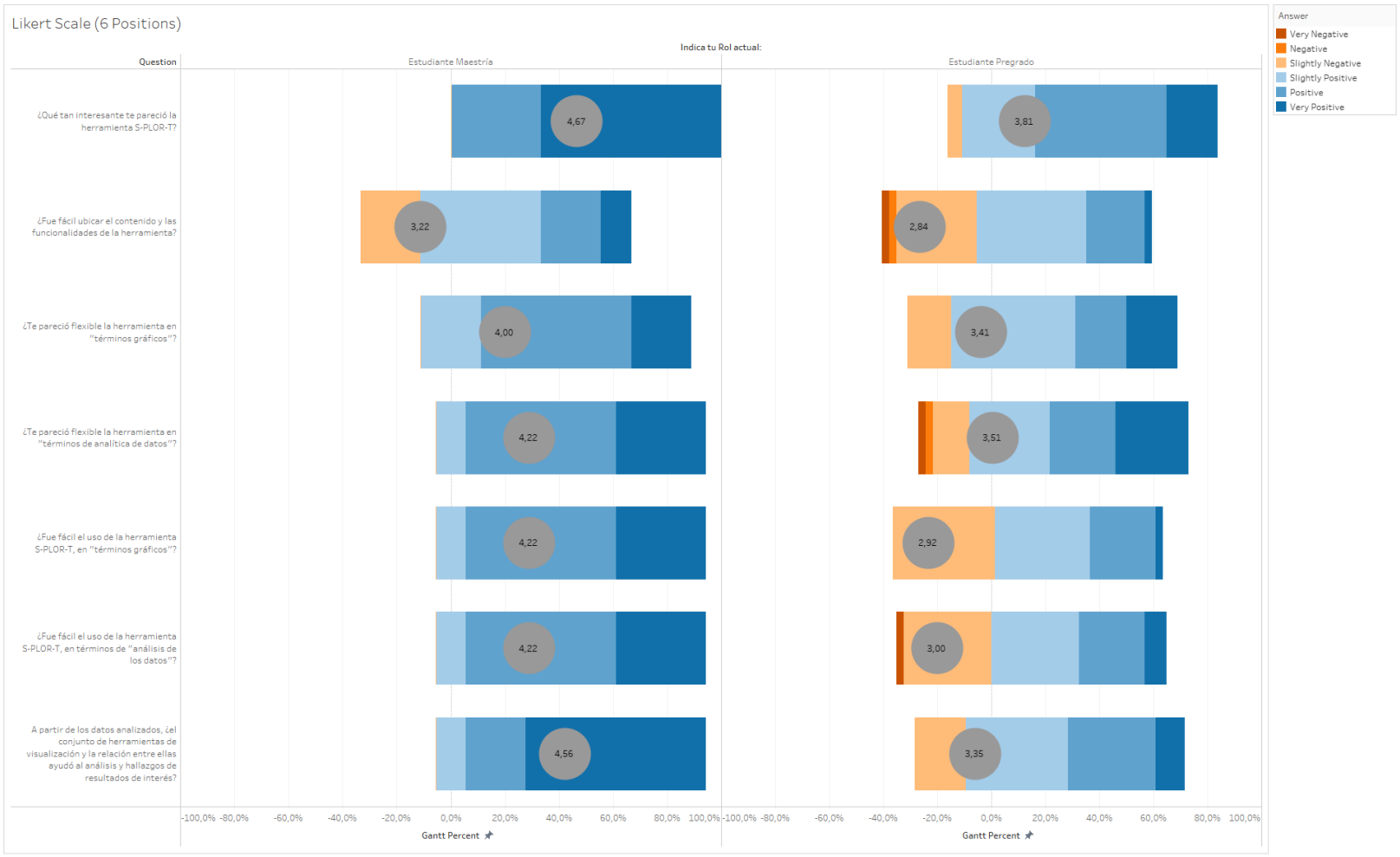


Figure 41 Likert Scale Analysis

Considering the previous plot, representing a Likert Scale for 6 positions, differencing both participating groups of the study, can be seen how the performance for each question previously described went. In this order of ideas, is evident that the participants who are considered as *experts* or that have somehow experience on analytics and having knowledge about the spatiotemporal domain, had a much better performance in general terms than the undergraduate students. Notwithstanding the last, it should be considered that in terms of facility on graphics terms, based on where to find the analysis desired, must be considered due to the will of the tool to be friendly to the user. Besides this, the

performance of the tool through the perception of the users (graduates - *experts*) went in a very positive way.

Furthermore, from the Likert plot above, can be also affirmed that even though the tool is thought to be somehow intuitive to the users, the target population should be not to all both expert and non-expert people, but to the ones who have a greater experience on managing information, knowing about alternatives of visual statistics and interpreting results.

For more details on the Quality Assessment of the tool, refer to the Annex in the Section 12.3 : "Assessment of Quality of the Tool: S-PLOR-T"

8. ANALYSIS OF RESULTS

From the results presented and described in the previous chapter, it is evident in general terms how the participation groups of the study differ. As it was well described, the exposed results and the existing difference between both groups measured can be the product of several factors. Perhaps this may be a consequence of the expertise of each participant regarding the domain in which the user tests were centered or on how each knows about data analytics with its multiple alternatives of analysis, represented some of them in the tool posed (e.g. multiple descriptive statistics).

Similarly, regarding the time constraint, was always tried to maintain the experimental methodology for all the user tests. However, for example, due to the time provided for the Civil Engineering Undergraduate course, the quality of the results is the implication of the time that was provided for the development of the tests. However, in general terms, the level of interest in the tool is positive for most of the participants, and in terms of the level of completeness and flexibility of this tool from different perspectives, the proportion of the population that considers that the tool does not meet the graphic and usage expectations (tool objectives), was minimal.

On the other hand, it is interesting to observe how, as the participants were completing the proposed challenges, it ended up being reflected in the invested times by the development of these, evidencing improvements (time reduction), in general terms. The above shows that the participants found their use more user-friendly as they became familiar with the tool as they were using it and its various functionalities. In colloquial terms, it can be emphasized that for any tool, for user interaction, such as happened with S-PLOR-T, will require time of knowledge and context of this, that although all the participants were introduced to the functionalities in their greater generality of the proposed tool: S-PLOR-T, until the user dives into the tool, they will find it easier to use correctly. This in a much more detailed way seen with the users considered as experts of the SUR Research Group, in comparison with the Undergraduate students of the Civil Engineering Transport Systems course at the Universidad de los Andes. The above, since the participants indicated a greater versatility to face new proposed challenges, each one with a higher level of graphic difficulty and analysis.

In addition, following the results presented in Chapter 7, particularly in the subsection 7.2, it is observed that the tool and study was well received by the participants in a high proportion, indicating that the tool was interesting and satisfied with the graphic objectives and analysis.

Likewise, recapitulating the results presented in the chapter above, it is observed that the participating users indicated having understood the correct use of the tool for analysis either little or totally, considering criteria of completeness versus correctness for each of the challenges posed.

9. CONCLUSIONS

On the first hand, the web tool designed meets the main objective of being a facilitator for exploratory data analysis with spatiotemporal characteristics in a flexible and user-friendly manner, with a contextualized approach towards Visual Analytics. The latter, based on the results obtained from the interaction of all the participating users in the study with this one, beyond the experimental results that were previously developed and described.

On the other hand, it is concluded that the challenges posed with their varied level of complexity, presented an accurate approach and well received by the participating population. In general terms, those participants of the study carried out the user evaluation voluntarily, apart from the fact that they were all told of their freedom to participate or withdraw at any time during the process.

Additionally, compiling the results of the Chapter above, it is important to emphasize again that all the results differ among the participating groups, which allows the inferring that the level of maturity in data analytics is fundamental to overcome the learning barrier of the tool. The above is justified by the results obtained by the users through the challenge, obtaining the expert users (graduates) much more positive results in terms of correctness than the undergraduates. Apart, considering the general results obtained, also allows inferring that there is a high relationship between the results obtained in the challenges (time spent, correctness, completeness, among others per challenge) per participant with their perception of the tool S-PLOR-T, as it happened comparatively between the two groups participating in the study and can be checked above graphically.

Based on the results and based on the analysis of them carried out in Chapter 7, in general terms and answering the research question posed in the subsection 2.3, is affirmative. This study showed that a web scope for the flexible and pleasant development of visual analytics within the linked paradigm of what, when and where is a good alternative for users who require easy accessibility to agile exploratory analysis. The latter reflects that the tool offers the user alternatives of analysis from different perspectives

that makes the analysis accurate, before the correct use of the functionalities (which also has implication in the knowledge of the data to be used).

To finish with, as was evident in comments made by several participant users, they indicated that S-PLOR-T facilitates the time that is spent in the analysis in comparison with other tools in the market, for its accessibility as being Web and its complete synchronization between several windows (containers). However, they emphasized the prior knowledge of the data for analytics and the tool itself, even if the user was contextualized in the domain of the data and the generalities of the tool, and a guide with the detail of the tool was previously provided. and its functionalities, the description of variables and specific attributes of the database used in the user tests were requested.

Therefore, the tool proposed: S-PLOR-T has great opportunities for being a great web approach for data analytics with such characteristics that permits the analysis through the main pyramid of the paradigm of the WH* questions, as described in Chapter 4. Moreover, once the tool is adapted in graphic terms and analysis from all the comments received during the user tests, it is intended to be much more user-friendly, scalable and efficient, thus being a better facilitator tool for exploratory data analytics with spatiotemporal features.

10. FUTURE WORK

After having carried out this study and having designed and implemented the proposed solution tool: S-PLOR-T, a broad horizon is set for future work that involves the tool, in terms of analysis and opportunities that come with it.

In this study, a first prototype was evaluated that allows to develop a synchronization of diverse analyzes in the frame of the questions within the paradigm of WHAT, WHEN and WHERE. The above, considering limitations in terms of space (dependent calculations of the browser) and specific characteristics of analysis of any kind. That is, for a future study, apart from the tools proposed for analysis in any of the defined fronts, it is expected to continue feeding the tool with additional analytical features that allow greater clarity and certainty in the results of questions of interest about the data, as considering the proposal as a Toolbox for Analytics.

On the other hand, it is expected that a future study may have an independent execution of the browser, which is equivalent to the use of relational databases, in a first state to be followed, so as not to have space / memory restrictions. That is, letting the browser draw the queries and filters done in a database, so that it could manage bigger files than the ones supported by the browser.

A future study could include the evaluation of usability with challenges from different perspectives of analysis following levels of complexity according to a sample population better defined or more specific. The latter, with the purpose of quantifying the success (correctness) of the participants by challenge, as it was developed in the present study, to guarantee a more rapid familiarization with the tool in a much more uniform way. That is, this study could be carried out again, modifying the challenges posed for this purpose of prototyping. This, seeking to standardize the requirement and the number of challenges or tasks that are proposed in general terms of joint interaction between views, thus achieving that this variable has a minimal influence on future results.

As can be seen, there are several possibilities to extend this project, the current study being a satisfactory pilot test that shows that S-PLOR-T fulfills the purpose for which it was proposed, designed and implemented. The tool, being a box of alternatives for visual analytics, entails opportunities for extension not only in terms of additional content for this purpose, but also in terms of the methodology used, since the analyzes carried out in the tool intend to follow such Flexibility, that each user find answers to analysis from different perspectives. Therefore, the project has such opportunities for modification and improvements to the scheme initially proposed and executed for this pilot study.

11. BIBLIOGRAPHIC REFERENCES

[1] Batini, C., Barone, D., Cabitza, F., & Grega, S. (Febrero de 2011). A DATA QUALITY METHODOLOGY FOR HETEROGENEOUS DATA. *International Journal of Database Management Systems, 3*(1).

[2] Guerra Gómez, J. A., Ortiz Román, J. C., & Murillo Castillo, J. G. (2018). Navio: a visualization widget for summarizing, exploring and navigating large multivariate datasets [En revisión]. *Infovis.*

[3] Munzner, T., & Maguire, E. (2015). *Visualization Analysis & Design.* FL: Taylor & Francis Group. Obtenido de https://bit.ly/2FHJJxE

[4] Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy:What Data Quality Means to Data Consumers. (T. &. Ltd., Ed.) *Journal of Management Information Systems, 12*(4), 5-33.

[5] Doraiswamy, H., Tzirita Zacharatou, E., Miranda, F., Lage, M., Ailamaki, A., Silva, C. T., & Freire, J. (2018, May). Interactive Visual Exploration of Spatiotemporal Urban Data Sets using Urbane. In *Proceedings of the 2018 International Conference on Management of Data* (pp. 1693-1696). ACM.

[6] Chen, W., Huang, Z., Wu, F., Zhu, M., Guan, H., & Maciejewski, R. (2018). Vaud: A visual analysis approach for exploring spatiotemporal urban data. *IEEE Transactions on Visualization & Computer Graphics*, (9), 2636-2648.

[7] Andrienko, N., Andrienko, G., & Gatalsky, P. (2003). Exploratory spatiotemporal visualization: An analytical review. Journal of Visual Languages & Computing, 14(6), 503-541. doi:10.1016/s1045-926x(03)00046-6

[8] Burigat, S., & Chittaro, L. (2005). Visualizing the results of interactive queries for geographic data on mobile devices. Proceedings of the 2005 International Workshop on Geographic Information Systems - GIS 05. doi:10.1145/1097064.1097103

[9] Andrienko, N., & Andrienko, G. (2004). Interactive visual tools to explore spatiotemporal variation. Proceedings of the Working Conference on Advanced Visual Interfaces - AVI 04. doi:10.1145/989863.989940

[10] Fernandez-Prieto, D., Naranjo-Valero, C., Hernandez, J. T., & Hagen, H. (2017). STRAD Wheel: Web-Based Library for Visualizing Temporal Data. IEEE Computer Graphics and Applications, 37(2), 99-105. doi:10.1109/mcg.2017.27

[11] Peña-Bastidas, J., Guzmán-García, L. A., & Carrasco, J. A. (2018). Linking a Walking and a Land Use Model to Support Expansion Policies in Cities: A Case Study in Bogotá.

[12] Landesberger, T. von, Bremm, S., Andrienko, N., Andrienko, G., & Tekusová, M. (2012). Visual Analytics Methods for Categoric Spatiotemporal Data. IEEE, doi:10.1109/VAST.2012.6400553

[13] Segura, H. (2014) Spatiotemporal Visualizer: Online tool to visualize trajectory data using a time-window. Lisbon, Portugal.

[14] Senaratne, H., Mueller, M., Behrisch, M., Lalanne, F., Bustos-Jimenez, J., Schneidewind, J, Keim, D., & Schreck, T. (2017). Urban Mobility Analysis With Mobile NetworkData: A Visual Analytics Approach. IEEE Transactions on Intelligent Transportation Systems. doi:10.1109/TITS.2017.2727281

[15] Emberger Günter, Pfaffenbichler Paul, Riedl Leopold (2010). MARS meets ANIMAP: Interlinking the Model MARS with dynamic Internet Cartography. doi:10.4113/jom.2010.1079

[16] LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big Data, Analytics and the Path From Insights to Value. MITSloan Management Review. Research Feature. Vol.52 No.2.

12. ANNEXES

12.1. Usability Tool Guide

Universidad de los Andes

Guía Prueba de Usabilidad – Herramienta: S-PLOR-T

Marco de Investigación: Tesis - Maestría en Ingeniería de Información

Investigador: Miguel Alfonso Feijóo García

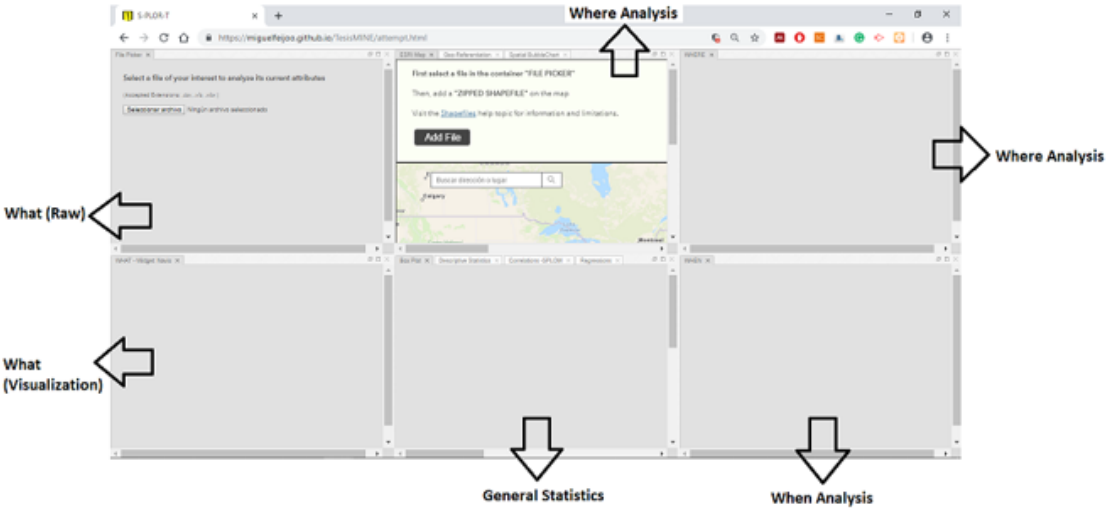
Bogotá, D.C., 2019

1. La herramienta: S-PLOR-T

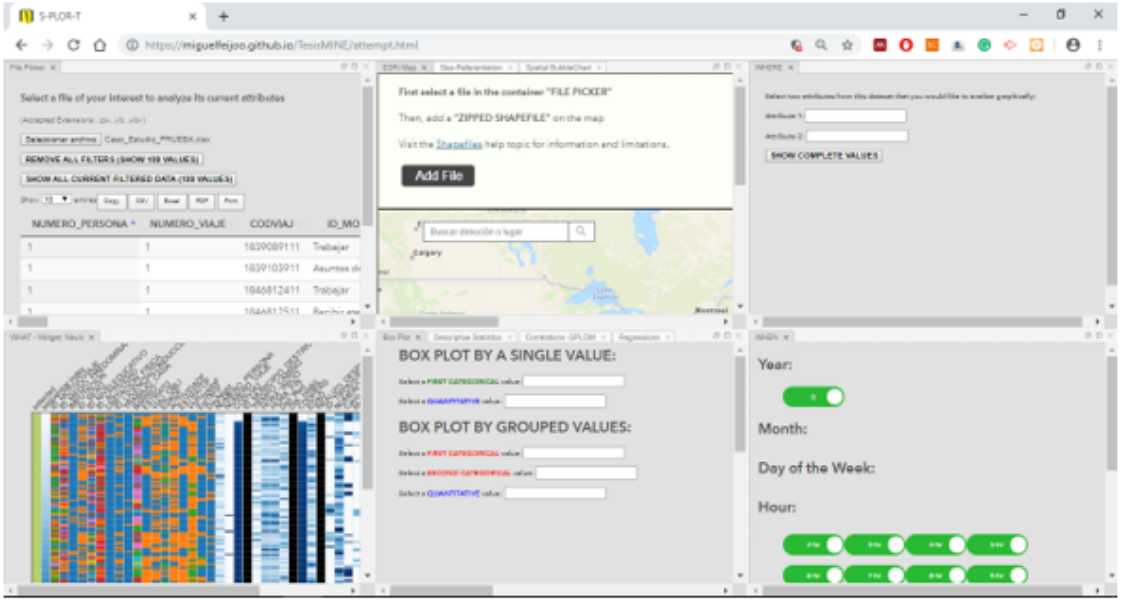
La herramienta: S-PLOR-T es una propuesta WEB para análisis exploratorio de datos con características espacio temporales de manera integrada, para considerar el paradigma conjunto del paradigma Wh* (WHAT-WHEN-WHERE). Se pretende de la herramienta garantizar una flexibilidad a la medida del usuario en el marco gráfico/interactivo, al igual que en el marco de análisis de información, desde diferentes puntos de vista mediante vistas enlazadas y sincronizadas.

1.1. Generalidades

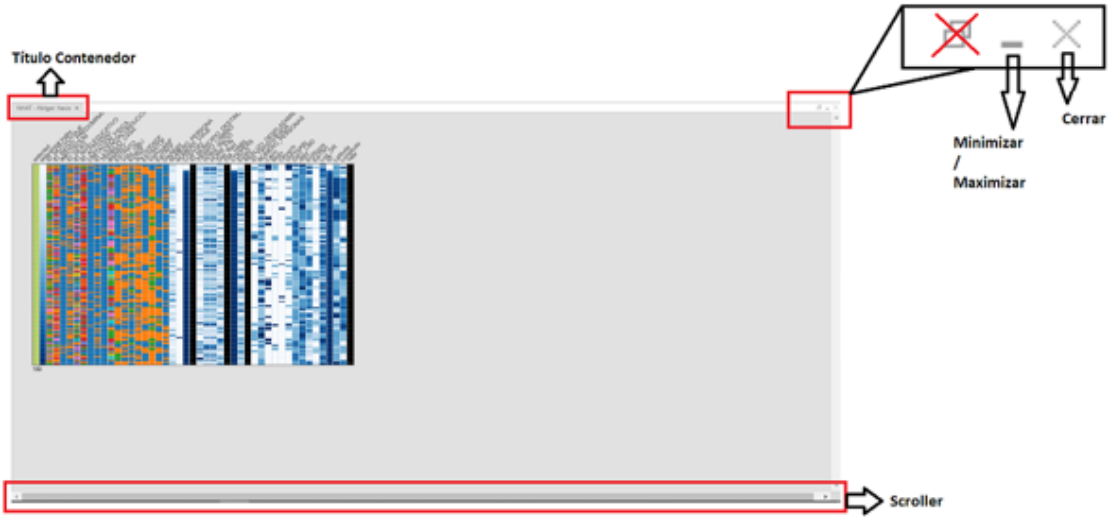
Ingresando al sitio: https://bit.ly/2ZluYqr, se puede acceder al prototipo de la herramienta planteada. Una vez ingresado a la URL, se encuentran seis (6) diferentes módulos que permiten la interacción en varias perspectivas.



Ahora bien, cada módulo es la representación de la visualización en uno de los paradigmas WH*. Esta vista en términos de "Contenedores" enlazados, enmarcados como ventanas.



A continuación, se presenta la relación de componentes por "Contenedor":



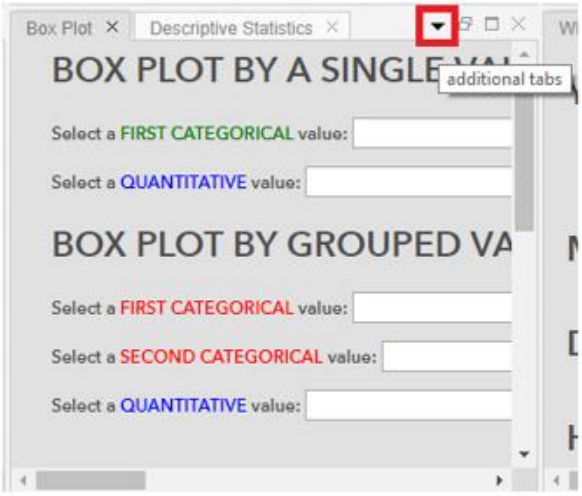
Mediante la imagen previa, es posible apreciar que cada uno de los contenedores tiene los elementos descritos, como:

- **Scroller**: Tanto horizontal como verticalmente.

- **Minimizar/Maximizar**: Cada contenedor se permite agrandar o achiquitar.
- **Cerrar**: Si el usuario lo desea, puede cerrar el contenedor que no le plazca ver en el instante del análisis.

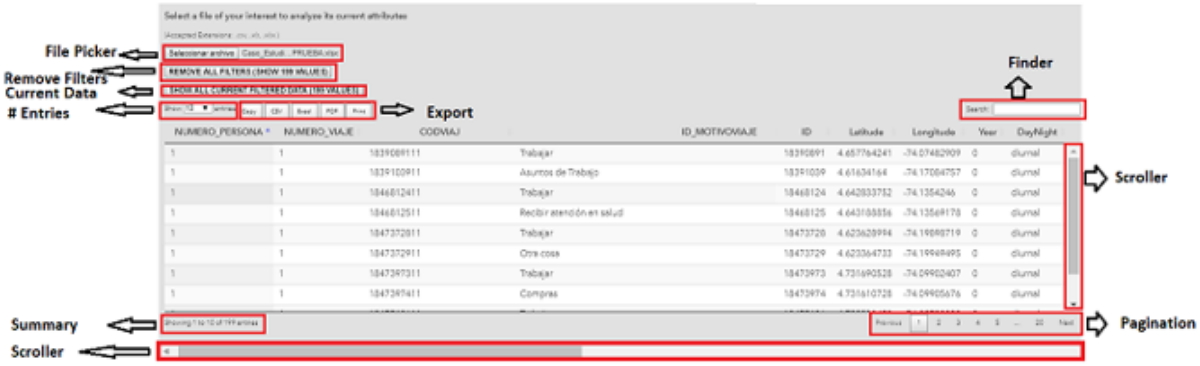
De igual forma, habrá contenedores que tienen de forma "stacked" en términos de pestañas para diferentes perspectivas dentro del dicho análisis. Por ejemplo, para los términos estadísticos:

Si se encuentra encogido el contenedor, se da a entender una mayor cantidad pestañas que las desplegadas con una flecha que indica las adicionales, como se muestra a continuación:



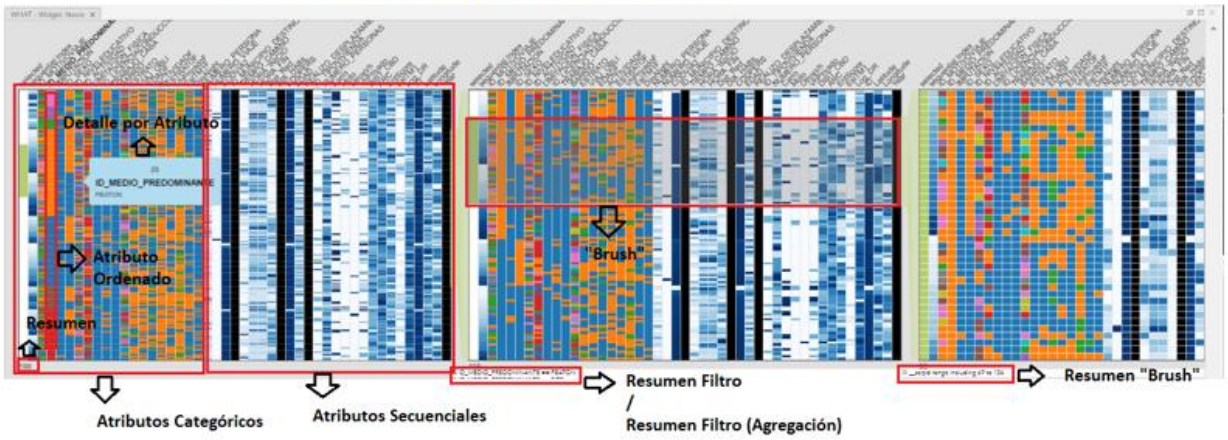
1.2. Interacción

1.2.1. Contenedor: File Picker



En este contenedor, es posible ver al detalle en texto los registros cargados/filtrados al interés del usuario. Contiene atributos como selección del archivo de análisis, que despliega la tabla con los atributos en crudo, el resumen desplegado distribuido mediante una paginación y por número de entradas, un buscador, y manejo de filtros o de la totalidad de la información a discreción de los intereses del usuario. Mediante la herramienta, es posible tener una forma de exportación en varios formatos para tener un resumen de la información general, seleccionada/filtrada.

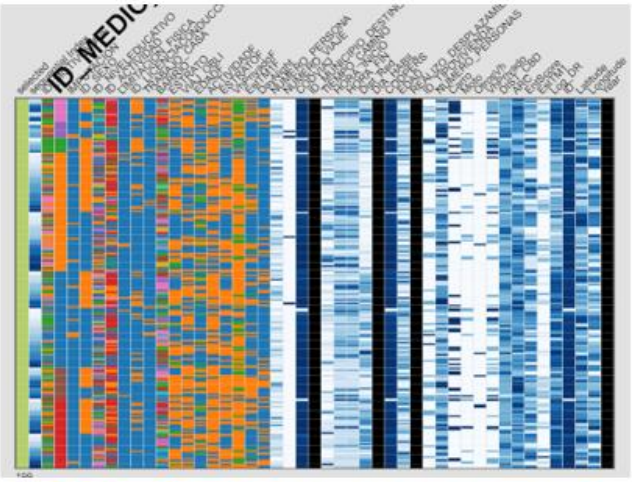
1.2.2. Contenedor: WHAT- Widget NAVIO



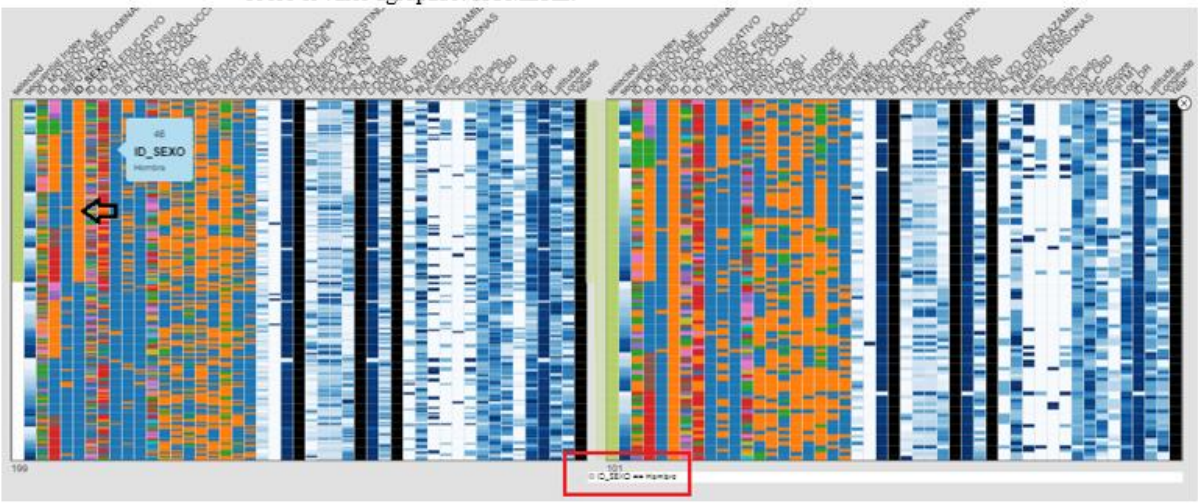
Este contenedor, representa el QUÉ (WHAT) dándole una noción al usuario de toda la información cargada. En la imagen previa se pueden apreciar los componentes de la visualización.

Para interactuar correctamente con la visualización:

1) En la parte superior, se encuentran todos los atributos (Categóricos y Secuenciales) distribuidos de forma que se puedan interpretar de la forma:
   a. **Atributos Categóricos**: Variación entre varios colores para grupos de valores.
   b. **Atributos Secuenciales**: Para un mismo color (por ejemplo, azul) se representan todos los valores como degradación de este (de mayor a menor, o viceversa).
2) **ORDENACIÓN**: Si se desea ordenar y/o agrupar por algún valor (práctico para valores categóricos), se busca el atributo en la parte superior (se va a ir sombreando) y se hace "click". De esta forma, se agruparán los valores de interés, como indica la imagen previa en **ATRIBUTO ORDENADO.**



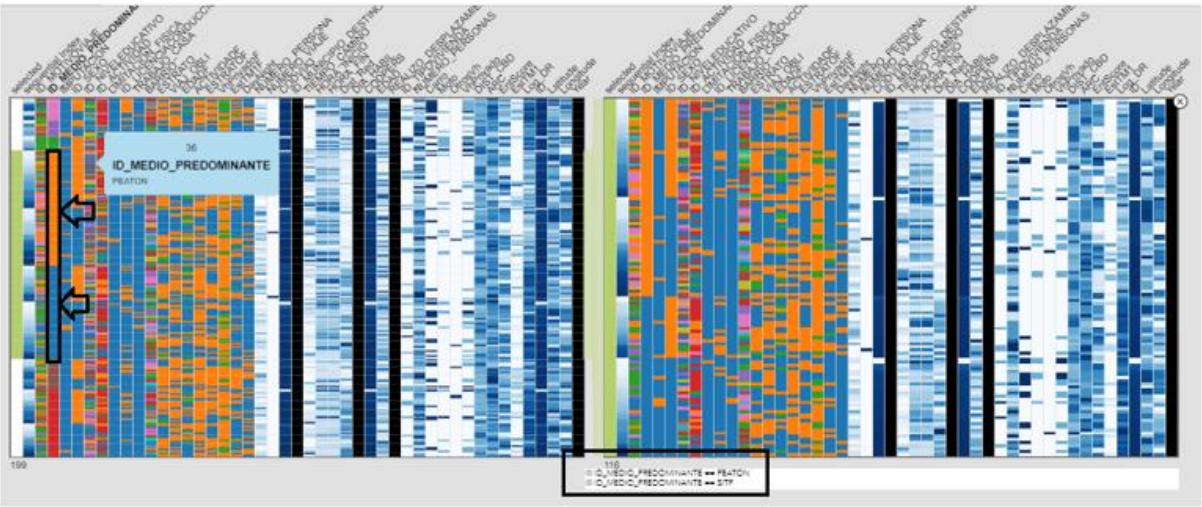
3) **FILTRO**: Para filtrar por **UN** solo valor, por ejemplo, HOMBRES, simplemente se requiere agrupar por ID_SEXO como se indicó previamente, y en seguida hacer ***click*** sobre el valor agrupado: HOMBRE.



De esta forma, se realiza el filtro sobre la totalidad de los datos desplegados en las otras vistas, pasando de 199 datos (para este ejemplo), a 101 que sólo son HOMBRES.

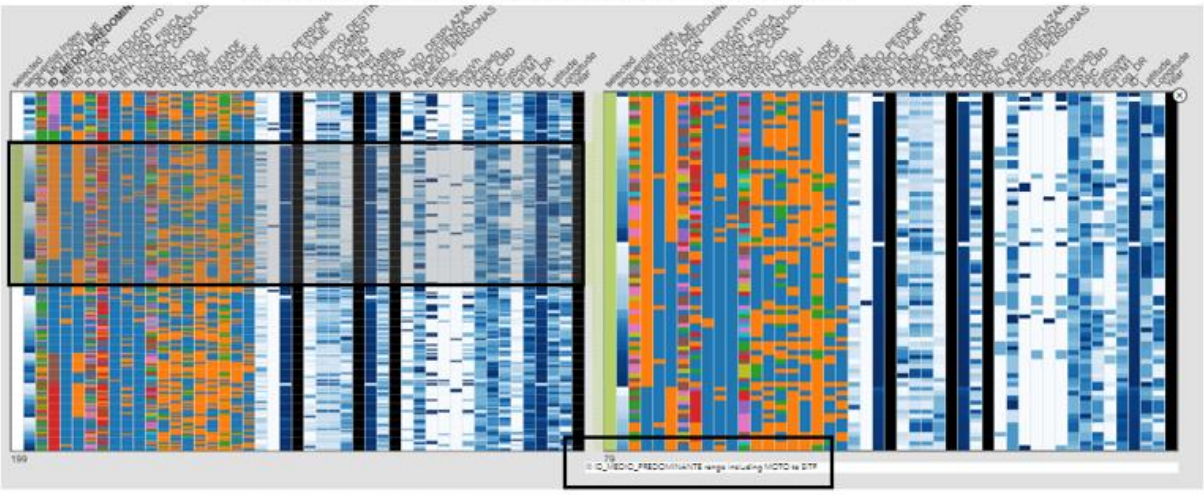
7

Asumiendo que se requiera una **AGREGACIÓN** sobre un filtro particular, se realiza el siguiente procedimiento. Por ejemplo, si se requiere de un filtro que indique ID_MEDIO_PREDOMINANTE (Medio de transporte Predominante) como PEATÓN y SITP, dicha agregación se puede realizar oprimiendo la tecla **SHIFT** mientras se realiza ***click*** sobre los dos valores (de un mismo atributo) de interés. En este caso, SHIFT → PEATON + SITP.



El resultado de la agregación se verá representada en el resultante como dos listas tal y como se indica en la imagen previa, en la parte inferior.

De igual forma, es posible realizar selección POR BARRIDO (BRUSHING) lo que permite al usuario seleccionar y filtrar a su discreción los valores a su medida. Para ello, manteniendo el ***click sostenido***, se realiza la selección de interés.



1.2.3. Contenedor: Statistics

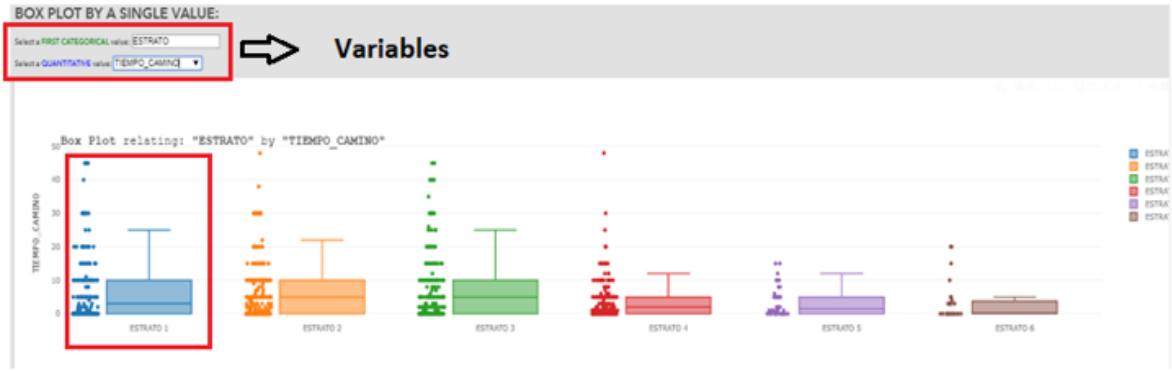
Este contenedor, hace énfasis en una muestra de análisis estadístico que se puede realizar con la información suministrada por el usuario, de manera comparativa y descriptiva por variable o conjunto de variables.

1.2.3.1. Contenedor: Box Plot

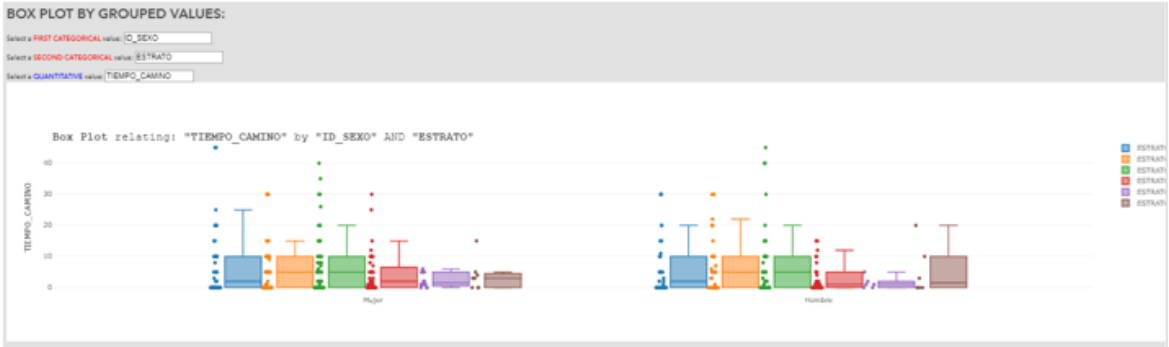
Análisis de Box and Whiskers (Cajas y Bigotes). Análisis estadístico para determinación de valores atípicos, valores dentro cuartiles o del rango Inter cuartil general.

Hay dos aproximaciones propuestas, que no necesariamente son las únicas:

- **ÚNICA variable categórica vs. Variable secuencial**

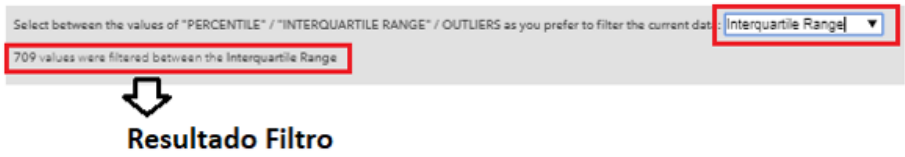


- **Variables Categóricas Agrupadas vs. Variable secuencial**



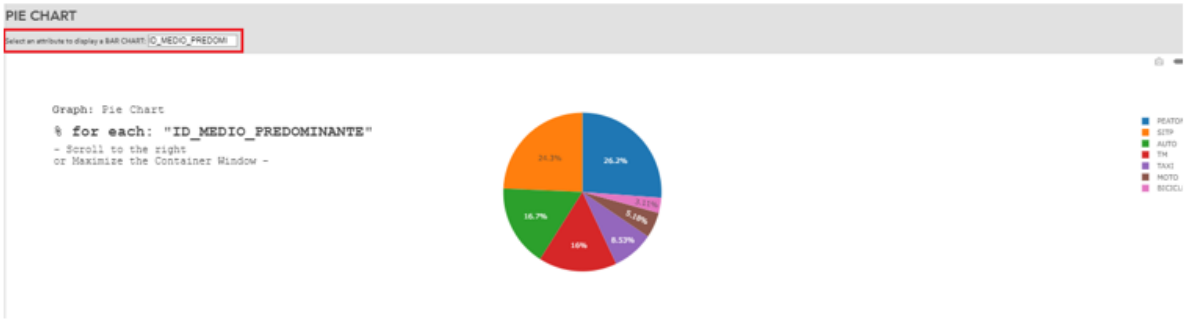
Para ambos casos, es posible realizar selección sobre alguna caja de interés, haciendo ***click*** sobre ella. Por ejemplo, si tengo un boxplot como el primero presentado (única variable), y requiero saber cuáles registros son los que hacen parte del ESTRATO 1, simplemente hago ***click*** sobre la caja y el resto de las vistas automáticamente se filtran a los registros restantes. Por otro lado, si requiero saber cuáles son los registros del ESTRATO 1 pero adicionalmente MUJERES y tengo la segunda representación de boxplot (agrupada por dos variables), hago ***click*** sobre la caja de ESTRATO 1 sólo en la agrupación de MUJERES.

Adicionalmente, si requiero tener noción de cuáles son los valores atípicos, o aquellos que hacen únicamente parte del rango inter-cuartil, en la lista de selección inferior, se puede realizar dicha selección:



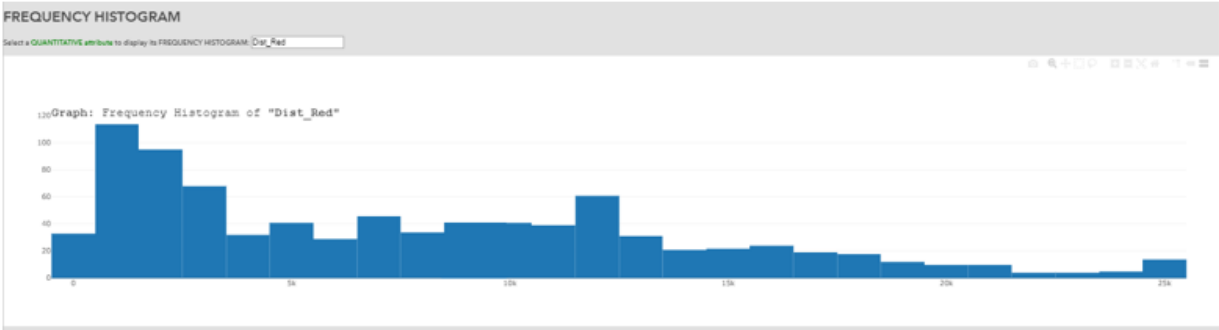
1.2.3.2. Contenedor: Descriptive Analysis

1.2.3.2.1. Pie Chart

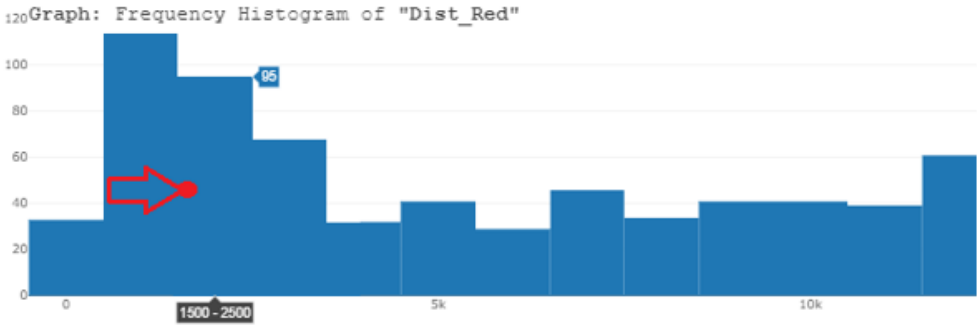


Este tipo de análisis representa para una variable particular una distribución en términos PORCENTUALES. Aparte de dar una noción sobre dicha distribución, el usuario también puede **FILTRAR** la totalidad de los registros a algún atributo de interés. Por ejemplo, saber cuáles registros son los que efectivamente utilizan como medio de transporte predominante: AUTO, el cual representa un 16.7% de la totalidad.

1.2.3.2.2. Frequency Histogram

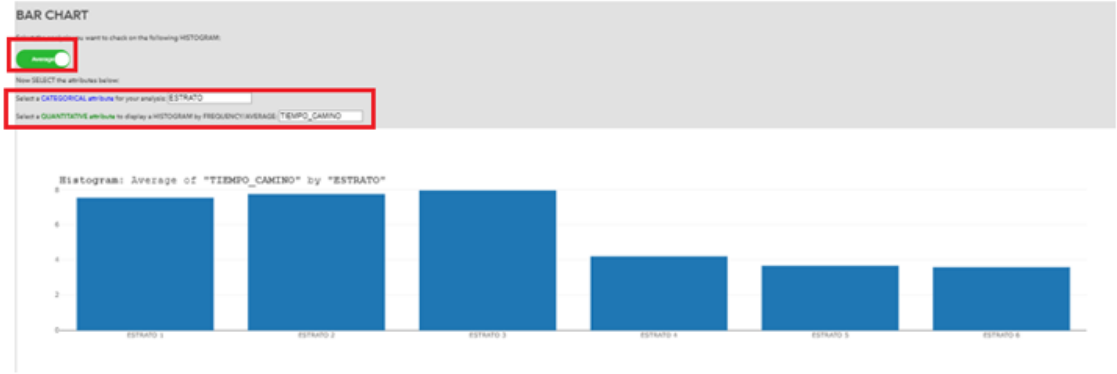


El presente, es un análisis de frecuencia. Lo anterior, implica que mediante el histograma se va a contar el número de registros que se encuentan dentro de un rango. Por ejemplo, si se desea saber cuántos registros se encuentran en un rango particular, sólo se requiere ver el número resultante al poner el ***mouse*** encima de una barra.



Para el caso previo, se indicaría que hay 95 registros que tienen como DISTANCIA DE VIAJE RECORRIDA entre 1500 y 2500 metros.

1.2.3.2.3. Single Bar Chart

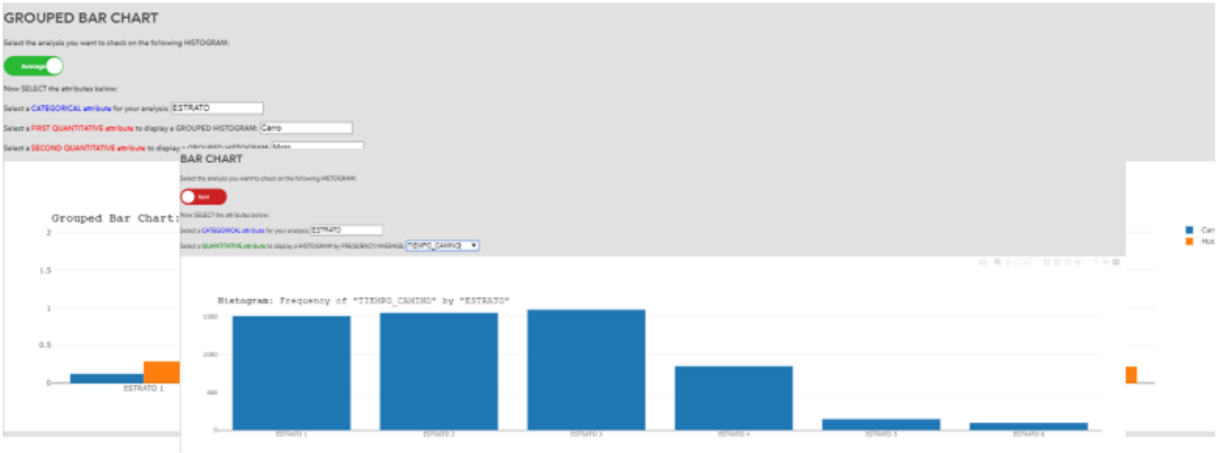


El análisis previo es un análisis que representa el **PROMEDIO** o la **SUMA** de una variable Secuencial y una variable Cualitativa. Se agrupan los valores a discreción del usuario y le da una noción general por grupo.

Para establecer si es PROMEDIO o SUMA, hay que tener en cuenta que el **BOTÓN VERDE** indica que se encuentra **ACTIVA** la opción de **PROMEDIO** (Average) y cuando se deselecciona y queda el **BOTÓN ROJO**, implica que se encuentra activa la opción de **SUMA** (Sum).

Aparte de dar una noción sobre la distribución desplegada, el usuario también puede **FILTRAR** la totalidad de los registros a algún atributo de interés.

1.2.3.2.4. Grouped Bar Chart

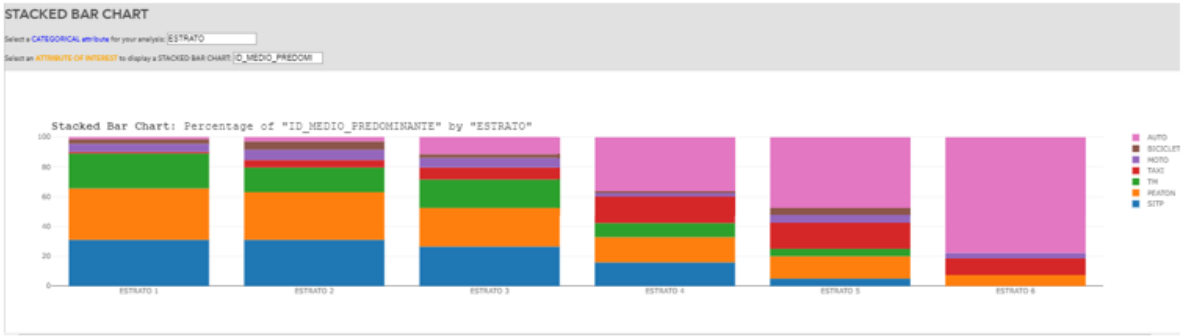


El análisis previo es un análisis que representa el **PROMEDIO** o la **SUMA** de dos variables Secuenciales y una variable Cualitativa. Se agrupan los valores a discreción del usuario y le da una noción general por grupo. Por ejemplo, el gráfico previo da una noción del PROMEDIO de Carros y de Motos que hay por ESTRATO.

Para establecer si es PROMEDIO o SUMA, hay que tener en cuenta que el **BOTÓN VERDE** indica que se encuentra **ACTIVA** la opción de **PROMEDIO** (Average) y cuando se deselecciona y queda el **BOTÓN ROJO**, implica que se encuentra activa la opción de **SUMA** (Sum).

Aparte de dar una noción sobre la distribución desplegada, el usuario también puede **FILTRAR** la totalidad de los registros a algún atributo de interés.

1.2.3.2.5. Stacked Barchart



El análisis previo es un análisis de relación de dos variables por conteo general, en representación PORCENTUAL de manera APILADA. Se agrupan los valores a discreción del usuario y le da una noción general por grupo.

Aparte de dar una noción sobre la distribución desplegada, el usuario también puede **FILTRAR** la totalidad de los registros a algún atributo de interés. Por ejemplo, filtrando únicamente los registros que tienen como MEDIO PREDOMINANTE: "AUTO" para ESTRATO 6.

1.2.3.3. Contenedor: Correlations -SPLOM

El presente análisis es la representación de una matriz de diagramas de dispersión de distintas variables para determinar si existen relaciones entre las variables del conjunto de datos. De esta forma, SPLOM es una forma efectiva de examinar cualquier correlación entre las variables.

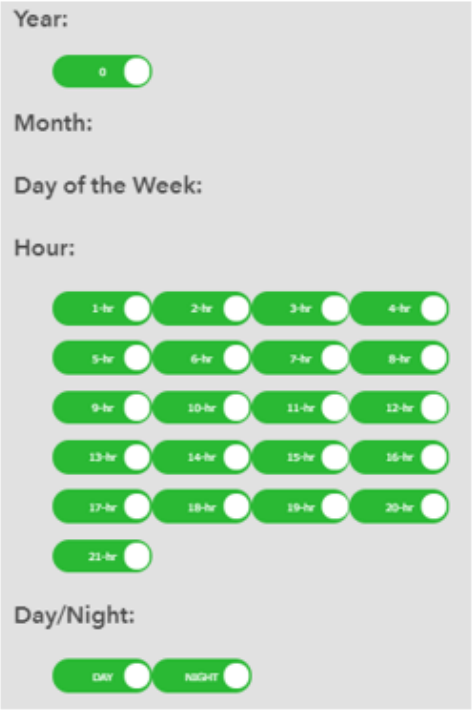
Haciendo "Brushing" o "Barrido" sobre cualquiera de los diagramas de dispersión, le permite tener al usuario una noción general de cuál es el comportamiento de entre las variantes de todas las variables para un grupo de interés. Para ello, sólo basta con hacer ***click*** sobre el diagrama de interés y se verá representado el comportamiento con el resto de las variables de dicha selección.

1.2.3.4. Contenedor: Regressions

El presente es un análisis que permite reconocer la dependencia y/o correlación entre variables aleatorias. Este análisis se realiza entre dos variables, particularmente, catalogado como REGRESIÓN SIMPLE. Da una noción al usuario de la interdependencia entre variables y su relación mediante pendiente de valores observados vs valores obtenidos, con un error particular de análisis.

1.2.4. Contenedor: WHEN

Contenedor especializado en el análisis TEMPORAL. A partir de la información suministrada, entre mayor sea el nivel de desagregación mayor cantidad de análisis temporal se podrá obtener. Para el presente ejemplo, se asumen sólo HORAS, que da una noción adicional, como si es de NOCHE o es de DÍA.



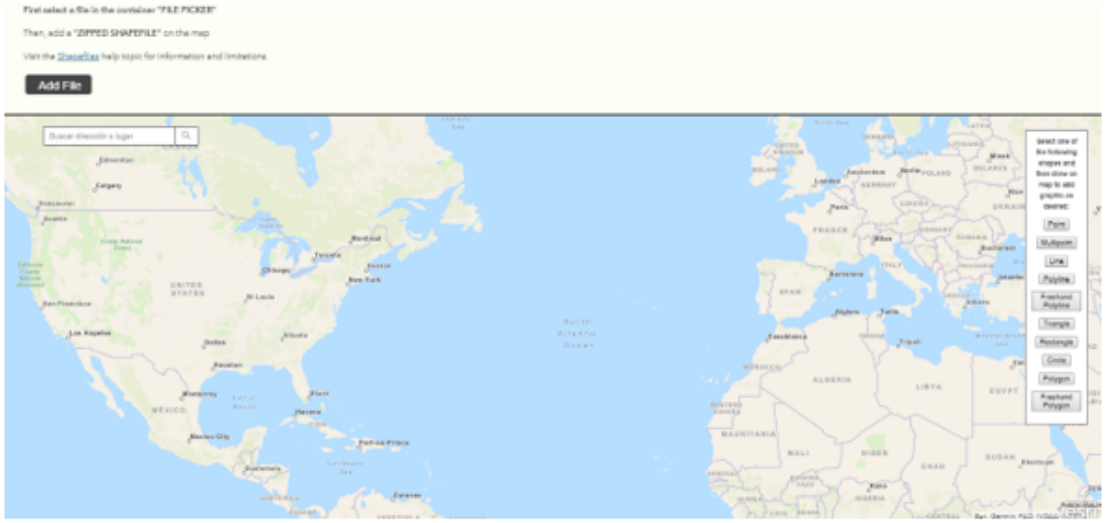
Como se mencionó previamente, este análisis permite establecer a discreción del usuario una selección de registros de interés. Por defecto, se encuentran todos los valores **ACTIVOS**, por lo que la interacción parte de la de **DESACTIVAR** valores o **REACTIVARLOS**. Para ello, solo es necesario hacer ***click*** sobre el valor que **NO** desea tener dentro de sus registros. Por ejemplo, si únicamente deseo tener los registros que hayan pasado de NOCHE, tendría que **DESACTIVAR** "DAY". Cuando se desactiva, queda el valor representado en **ROJO**:

1.2.5. Contenedor: Spatial Analysis

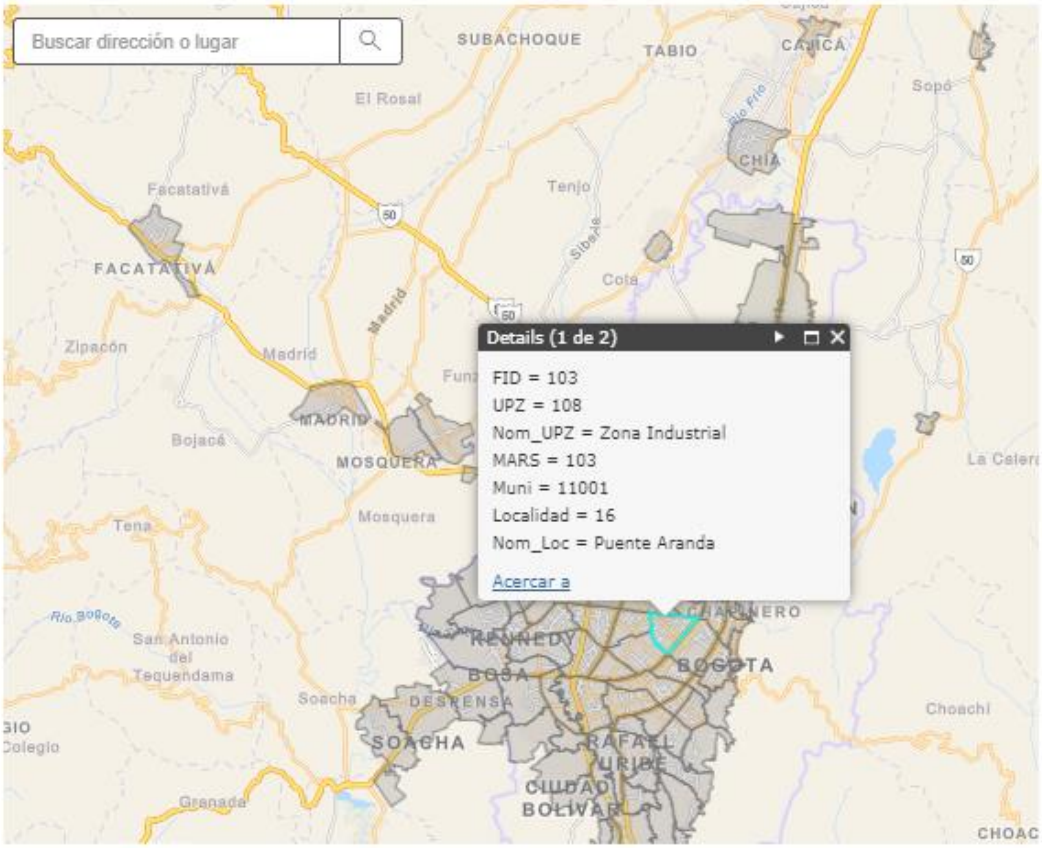
Este contenedor permite tener una noción espacial de la información, si los datos cargados la proveen. Hay representaciones gráficas como MAPA o como otras alternativas y/o variantes de representación espacial.

1.2.5.1. Contenedor: ESRI Map

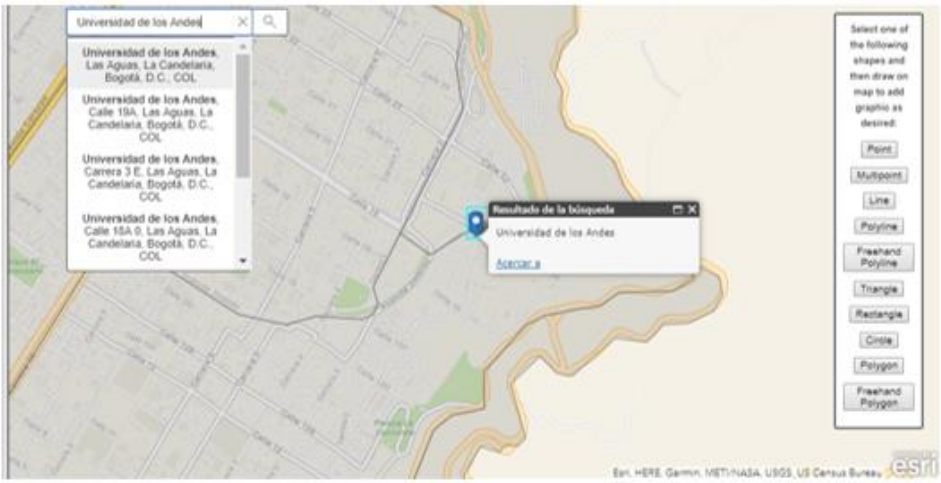
Contenedor útil si hay información con extensión SHAPEFILE y la información cargada como Excel o CSV provee un enlace directo con el shape.



Cargando el archivo con extensión ***.shp*** se logra obtener un resultado como el que se muestra a continuación:

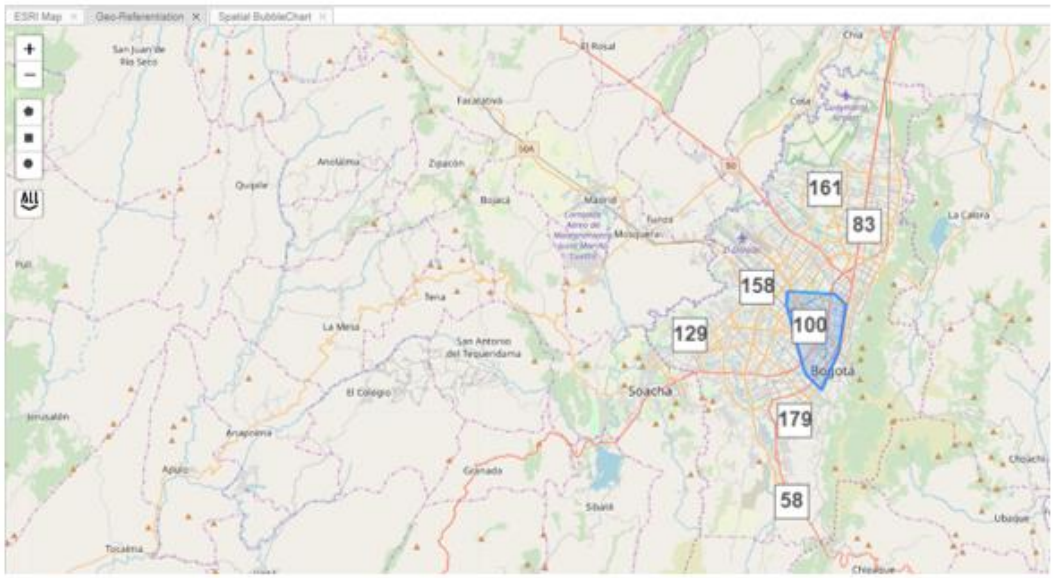


Si es el caso de filtrar por UPZ, como es el caso del ejemplo descrito en el archivo, basta con hacer ***click*** sobre la zona de interés y la totalidad de la información para todas las vistas se verá de igual forma alterada. En caso de no tener información particular sobre una UPZ, pero se desee saber información detallada de la misma, haciendo ***click*** sobre la zona desplegará el detalle de la misma.



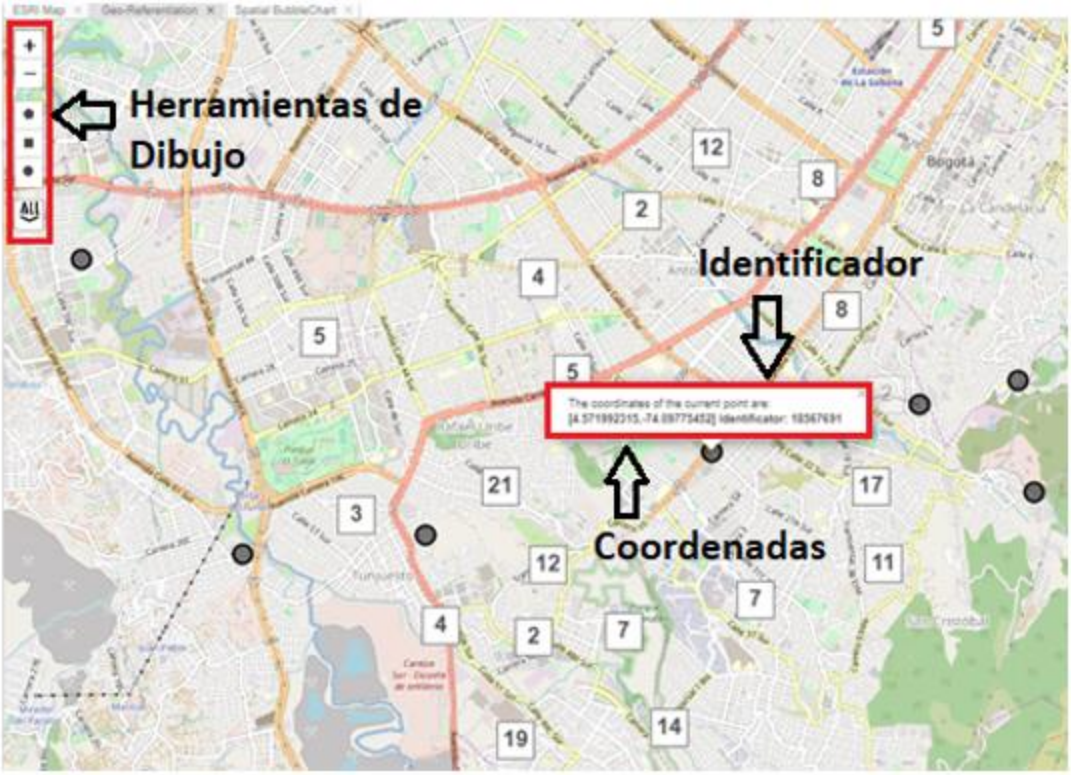
Ahora bien, el usuario puede buscar detalle en demanda sobre una o varias zonas de interés en el espacio, buscando direcciones o zonas de interés, los cuales aparecerán en el lienzo del mapa.

1.2.5.2. Contenedor: Geo-Referentiation



Esta pestaña del contenedor espacial es una representación de los puntos coordenados (x, y), si se encuentran disponibles en el archivo cargado por el usuario. Es decir, si hay alguna relación coordenada (latitud, longitud), esta información se verá reflejada en el lienzo del mapa como puntos aglomerados con ara de influencia. Dicha área de influencia, resaltada en AZUL, contendrá la cantidad de puntos definida en el recuadro. Para mayor detalle, el usuario

puede hacer zoom-in o zoom-out con el ***mouse***, o haciendo click sobre el recuadro de interés, para ver los puntos particulares de interés, como se muestra a continuación:



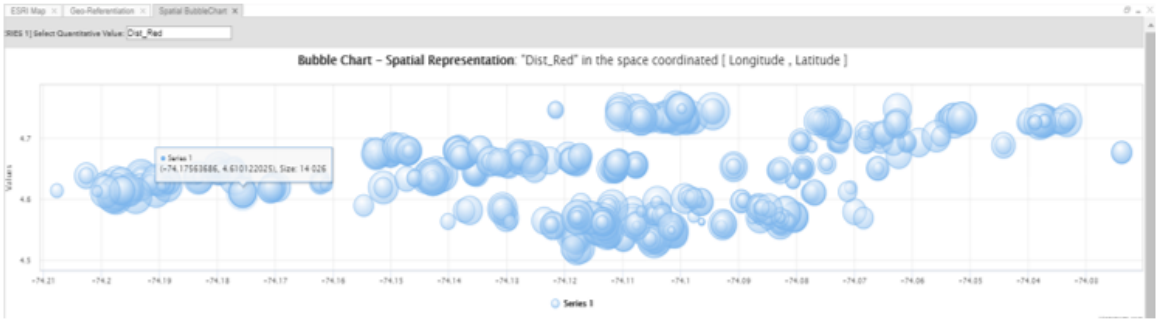
Como se puede apreciar, el usuario tiene la posibilidad de ver el detalle de cada punto coordenado. Como defecto, se despliega por punto información correspondiente a las COORDENADAS y al IDENTIFICADOR del punto de interés. Sin embargo, si hay relación de FECHA, esta también se desplegará sin problema.

El usuario, adicionalmente, tiene la posibilidad de FILTRAR a su demanda mediante POLILÍNEAS, ÁREAS: RECTANGULARES Y RADIALES, como lo indican las **HERRAMIENTAS DE DIBUJO**. Cuando el usuario lo desee, puede generar un **POLÍGONO**, haciendo ***click*** para generación de puntos hasta cuando se cierre la figura de interés. Los puntos coordenados dentro del polígono serán filtrados, dejando sólo los registros

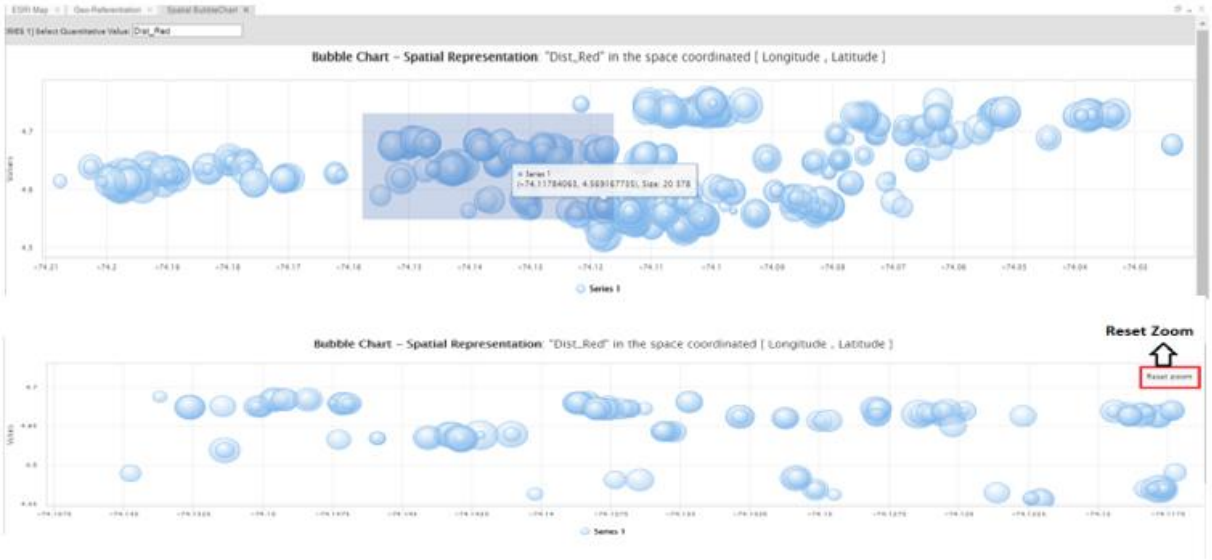
enmarcados por las líneas definidas en la figura. Por otro lado, se puede definir RADIO o Área, siguiendo la misma metodología previa, empezando la figura con un ***click*** y terminándola con otro ***click***.

1.2.5.3. Contenedor: Spatial Bubble Chart

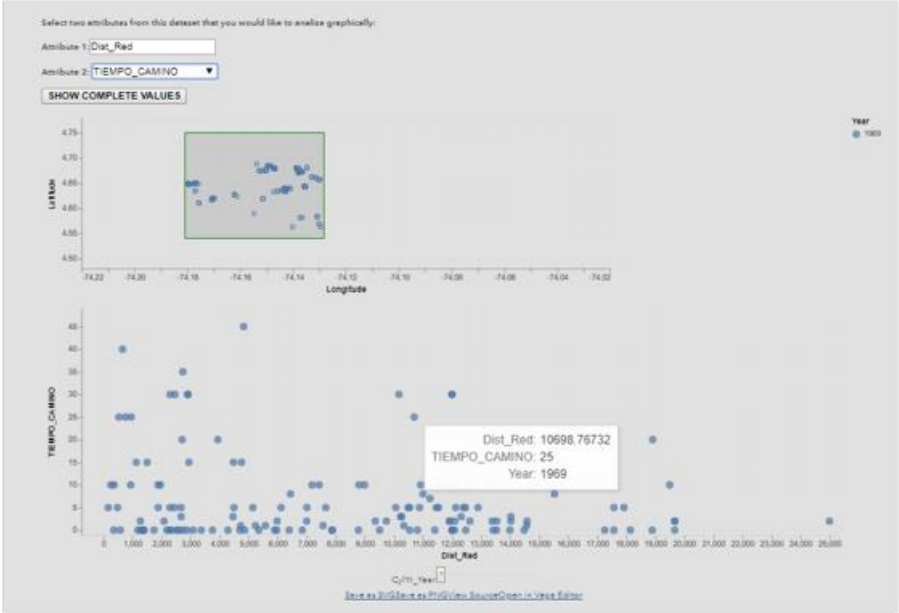
Esta pestaña del contenedor permite al usuario tener una noción del espacio de una forma diferente, mediante la distribución de puntos coordenados y el comportamiento de una variable de interés en el mismo. De esta forma, se despliega un gráfico con características espaciales, sin requerimiento o dependencia de un mapa, en donde en el punto (latitud, longitud) [x,y], se encuentra la representación de una variable por selección y discreción del usuario en términos de tamaño.



Si el usuario desea mayor detalle de una sección del gráfico, haciendo ***click sostenido*** puede hacer zoom-in. Para volver a la perspectiva gráfica inicial, es solamente hacer ***click*** sobre el botón de RESET ZOOM, ubicado en l parte superior derecha, como se indica a continuación:



1.2.6. Contenedor: WHERE



23

Este contenedor, al igual que el anteriormente descrito, permite al usuario tener una noción del espacio de una forma diferente, mediante la distribución de puntos coordenados y el comportamiento de una variable de interés en el mismo. De esta forma, se despliega un gráfico con características espaciales, sin requerimiento o dependencia de un mapa, en donde en el punto (latitud, longitud) [x, y], se encuentra la representación de la relación entre dos variables por selección y discreción del usuario. Para darle un mayor detalle al usuario, él puede establecer la información que desea ver desplegada, haciendo "brush" o "barrido" en el espacio coordenado (gráfica superior), viéndose alterada la información relacional (gráfica inferior). Haciendo ***click*** sobre un punto coordenado de interés, se proveerá al usuario detalle de dicha relación.

2. Ejemplo – Caso Reto Prueba

En la prueba de usabilidad propuesta, se deben resolver cinco (5) retos que permiten una interacción con varios elementos/componentes de la herramienta implementada. Los datos utilizados es una muestra pequeña de la ENCUESTA DE MOVILIDAD para el año 2015, realizada a unos hogares de la ciudad de Bogotá D.C.

A continuación, se encuentra un resumen de cada uno de los atributos usados, con su descripción y un ejemplo respectivo para un registro:

NOMBRE POST-PREPROCESAMIENTO	DESCRIPCIÓN	EJEMPLO
ID	Identificador de la Encuesta por Hogar	18475811
NUMERO_PERSONA	-	-
NUMERO_VIAJE	Número de Personas por Hogar	1
CODVIAJ	Número de Viajes por Encuestado	1847581111
ID_MOTIVOVIAJE	Motivo por el que el encuestado realizó el viaje	Estudiar
ID_MUNICIPIO_DESTINO	Municipio en el que el encuestado realizó el viaje (origen-destino)	1
TIEMPO_CAMINO	Tiempo Invertido en el viaje en minutos	20
HORA_INICIO	Hora en la que inició el viaje (En términos de 24 Horas)	9
HORA_FIN	Hora en la que terminó el viaje (En términos de 24 Horas)	10
Dist_Red	Distancia de Viaje recorrida por encuestado desde su origen hasta su destino (en metros)	11638.02773

NOMBRE POST-PREPROCESAMIENTO	DESCRIPCIÓN	EJEMPLO
ID_MEDIO_PREDOMINANTE	Medio de Transporte Predominante por Encuestado	SITP
LATITUD	Latitud del punto origen por encuestado/hogar	4,495027789
LONGITUD	Longitud del punto origen por encuestado/hogar	-74,11609858
IMPUTACION	-	-
CODPERS	Código que indica la persona encuestada	184758111
ID_SEXO	Indica el Género de la persona Encuestada	Mujer
EDAD	Indica la EDAD de la persona encuestada	55
ID_NIVELEDUCATIVO	Indica el nivel educativo de la persona encuestada	Primara Completa
ID_ACTIVIDAD	Indica la actividad a la que se dedica la persona encuestada	Estudia
LIMITACION_FISICA	Indica si la persona encuestada tiene limitación física o no	NO
ID_LICENCIACONDUCCION1	Indica si la persona encuestada tiene licencia	NO_LICENCIA
REALIZO_DESPLAZAMIENTO	Indica si la persona encuestada realizó desplazamiento	1
TRABAJO_CASA	Indica si la persona encuestada trabaja en su propio hogar o no	No
BARRIO	Indica el barrio en el que se encuentra el hogar encuestado	MODELIA
ID_TIPOVIVIENDA	Indica el tipo de la vivienda del hogar encuestado	4
NUMERO_PERSONAS	Indica el número de personas que hacen parte del hogar encuestado	3
ESTRATO	Indica el Estrato al que pertenece el hogar encuestado	ESTRATO 4
Carro	Indica el número de Carros que tiene el hogar encuestado	2
Moto	Indica el número de Motos que tiene el hogar encuestado	0
OtrosVh	Indica el número de otro tipo de vehículos que tiene el hogar encuestado	0

NOMBRE POST-PREPROCESAMIENTO	DESCRIPCIÓN	EJEMPLO
VhPrivado	Indica el número de vehículos privados que tiene el hogar encuestado	2
DIST_CBD	-	-
ARC	-	-
EntScore	-	-
EstTM1	Indica el número de estaciones de Transmilenio cerca del hogar encuestado	2
VIAJ_OBLI	Indica si el viaje del encuestado fue obligado o no	No_Oblig
EDADF	Rango de edades	(22,56]
ACTIVIDADF	Rango de actividades	Estudia/Trabaja
ESTRATOF	Rango de estratos	Alto
VhPrivadoF	Rango del número de vehículos privados	Dos o mas
EstTM1F	Rango de estaciones de Transmilenio cerca del hogar	Bajo
Log_DR	-	-
DayNight	Indica si el registro del viaje realizado por el encuestado se dio de DIA o de NOCHE	Diurnal
Year	Indica el/los año(s) por registro	0

Para descargar el archivo de prueba, has ***click*** en el siguiente enlace: **ENLACE ARCHIVO**

12.2. Usability Evaluation

EVALUACIÓN DE USABILIDAD

Herramienta: *S-PLOR-T*

Maestría en Ingeniería de Información
Universidad de los Andes
2019

---

*Autorizo al investigador **MIGUEL ALFONSO FEIJÓO GARCÍA**, estudiante de **MAESTRÍA EN INGENIERÍA DE INFORMACIÓN (MINE)** a obtener, tabular y analizar los datos proveídos en esta evaluación de usabilidad en el marco de investigación: **TESIS**, en la UNIVERSIDAD DE LOS ANDES – FACULTAD DE INGENIERÍA.*

[*Marca con una X tu respuesta*]

Sí Autorizo ☐ **No Autorizo** ☐

FIRMA: _____

Fecha: ____ / ____ / ____ ***(dd/mm/a)***

Hora: ____ : ____ ***(hh:mm)***

Nombre: _____

EVALUACIÓN

- **Reto No. 1**: ***"Identificar la frecuencia de las personas que recorren una distancia de red/viaje (Dist_Red) entre 3000 y 4999 metros (3 - 4.99 kilómetros), con MEDIO DE TRANSPORTE PREDOMINANTE (ID_MEDIO_PREDOMINANTE): PEATÓN y BICICLETA"***

1. Respuesta: _____ (*También es válido decir "No sé"*)
2. ¿Pudiste completar el reto propuesto en la herramienta S-PLOR-T? [*Marca con una X tu respuesta*]

Sí ☐ **No** ☐

3. Si tu respuesta a la pregunta anterior fue ***SÍ***, ¿fue fácil resolver el reto? [*Marca con una X tu respuesta*]

Sí ☐ **No** ☐

Cuéntanos el porqué de tu respuesta:

_____

_____

_____

4. Indícanos 3 aspectos (o menos, si prefieres) **POSITIVOS** respecto a la resolución de este reto:
   - **i.** _____
   - **ii.** _____
   - **iii.** _____

5. Indícanos 3 aspectos (o menos, si prefieres) **NEGATIVOS** o para **MEJORAR** respecto a la resolución de este reto:
   - **i.** _____
   - **ii.** _____
   - **iii.** _____

Tiempo tardado en el reto (*no importa si no se completó el reto*): ***minutos*** ***segundos***

¡Muchas gracias por tu colaboración!

- **Reto No. 2**: ***"Identificar cuál ESTRATO (ESTRATO) tiene como SUMA el mayor número de CARROS (Carros) y cuál el mayor número de MOTOS (Motos)."***

1. Respuesta: _____ (*También es válido decir "No sé"*)
   _____ (*También es válido decir "No sé"*)
2. ¿Pudiste completar el reto propuesto en la herramienta S-PLOR-T? [*Marca con una X tu respuesta*]

   Sí ☐ **No** ☐

3. Si tu respuesta a la pregunta anterior fue ***SÍ***, ¿fue fácil resolver el reto? [*Marca con una X tu respuesta*]

   Sí ☐ **No** ☐

   Cuéntanos el porqué de tu respuesta:
   _____
   _____
   _____

4. Indícanos 3 aspectos (o menos, si prefieres) **POSITIVOS** respecto a la resolución de este reto:
   i. _____
   _____
   ii. _____
   _____
   iii. _____
   _____

5. Indícanos 3 aspectos (o menos, si prefieres) **NEGATIVOS** o para **MEJORAR** respecto a la resolución de este reto:
   i. _____
   _____
   ii. _____
   _____
   iii. _____
   _____

Tiempo tardado en el reto (*no importa si no se completó el reto*): ***minutos*** ***segundos***

¡Muchas gracias por tu colaboración!

- **Reto No. 3**: ***"Identificar cuáles son los dos (2) MEDIOS DE TRANSPORTE PREDOMINANTES (ID_MEDIO_PREDOMINANTE) con mayor porcentaje para únicamente las MUJERES (ID_SEXO) que se encuentran en un rango de edad (22,56] años (EDADF), únicamente en las horas de la NOCHE (NIGHT) – (day/night). Ahora bien, d*el resultado de los filtros previos, indica en qué parte se concentra el mayor número de MUJERES transportándose de NOCHE en cualquier medio de transporte. Para responder esta pregunta, indica: NORTE, SUR, ORIENTE, OCCIDENTE."**

1. Respuesta: _____ *(También es válido decir "No sé")*
_____ *(También es válido decir "No sé")*
_____ *(También es válido decir "No sé")*

2. ¿Pudiste completar el reto propuesto en la herramienta S-PLOR-T? [*Marca con una X tu respuesta*]

Sí ☐ **No** ☐

3. Si tu respuesta a la pregunta anterior fue ***Sí***, ¿fue fácil resolver el reto? [*Marca con una X tu respuesta*]

Sí ☐ **No** ☐

Cuéntanos el porqué de tu respuesta:
_____
_____
_____

4. Indícanos 3 aspectos (o menos, si prefieres) **POSITIVOS** respecto a la resolución de este reto:

i. _____
_____

ii. _____
_____

iii. _____
_____

5. Indícanos 3 aspectos (o menos, si prefieres) **NEGATIVOS** o para **MEJORAR** respecto a la resolución de este reto:

i. _____
_____

ii. _____
_____

iii. _____
_____

Tiempo tardado en el reto (*no importa si no se completó el reto*): ***minutos*** ***segundos***

¡Muchas gracias por tu colaboración!

- **Reto No. 4**: ***"Indicar en qué BARRIO (BARRIO) se registra la mayor distancia de viaje (Dist_Red) en PROMEDIO, para únicamente los HOMBRES (ID_SEXO) con un rango de edad (EDADF) de (56,99] con MEDIO DE TRANSPORTE PREDOMINANTE (ID_MEDIO_PREDOMINANTE): PEATÓN."***

1. Respuesta: _____ (*También es válido decir "No sé"*)
2. ¿Pudiste completar el reto propuesto en la herramienta S-PLOR-T? [*Marca con una X tu respuesta*]

Sí ☐ **No** ☐

3. Si tu respuesta a la pregunta anterior fue ***SÍ***, ¿fue fácil resolver el reto? [*Marca con una X tu respuesta*]

Sí ☐ **No** ☐

Cuéntanos el porqué de tu respuesta:

_____

_____

_____

4. Indícanos 3 aspectos (o menos, si prefieres) **POSITIVOS** respecto a la resolución de este reto:
   - **i.** _____
   - **ii.** _____
   - **iii.** _____

5. Indícanos 3 aspectos (o menos, si prefieres) **NEGATIVOS** o para **MEJORAR** respecto a la resolución de este reto:
   - **i.** _____
   - **ii.** _____
   - **iii.** _____

Tiempo tardado en el reto (*no importa si no se completó el reto*): ***minutos*** ***segundos***

¡Muchas gracias por tu colaboración!

- **Reto No. 5**:
  - i. ***"Identificar el porcentaje de personas con un rango de edad (EDADF) de (11,22] años cuyo MEDIO DE TRANSPORTE PREDOMINANTE (ID_MEDIO_PREDOMINANTE) sea SITP. Indicar el número de personas resultantes del análisis"***
  - ii. ***Visualizar un BOX-PLOT donde se evidencie la distancia de viaje recorrida (Dist_Red) por ESTRATO (ESTRATO). Del análisis previo, identificar el número de personas que son del ESTRATO 2 (ESTRATO) que, adicionalmente, cumplen con los filtros previos.***
  - iii. ***Adicionalmente, indicar el porcentaje del NIVEL EDUCATIVO (ID_NIVELEDUCATIVO) que es mayor para únicamente HOMBRES (ID_SEXO)"***

1. Respuesta: **i.** _____ (*También es válido decir "No sé"*)
   ii. _____ (*También es válido decir "No sé"*)
   iii. _____ (*También es válido decir "No sé"*)
2. ¿Pudiste completar el reto propuesto en la herramienta S-PLOR-T? [*Marca con una X tu respuesta*]

   Sí ☐ **No** ☐

3. Si tu respuesta a la pregunta anterior fue ***SÍ***, ¿fue fácil resolver el reto? [*Marca con una X tu respuesta*]

   Sí ☐ **No** ☐

   Cuéntanos el porqué de tu respuesta:
   _____
   _____
   _____

4. Indícanos 3 aspectos (o menos, si prefieres) **POSITIVOS** respecto a la resolución de este reto:
   i. _____
   _____
   ii. _____
   _____
   iii. _____
   _____

5. Indícanos 3 aspectos (o menos, si prefieres) **NEGATIVOS** o para **MEJORAR** respecto a la resolución de este reto:
   i. _____
   _____
   ii. _____
   _____
   iii. _____
   _____

Tiempo tardado en el reto (*no importa si no se completó el reto*): ***minutos*** ***segundos***

¡Muchas gracias por tu colaboración!

12.3. Assessment of Quality of the Tool: S-PLOR-T

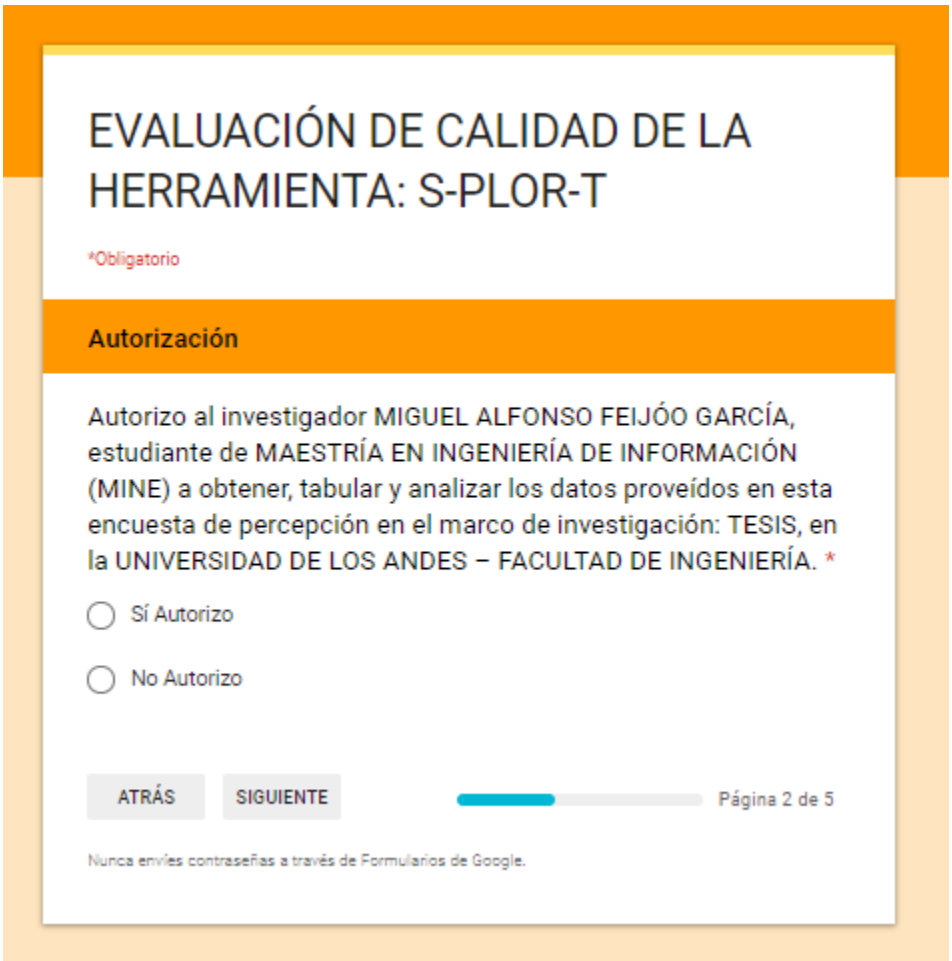
EVALUACIÓN DE CALIDAD DE LA HERRAMIENTA: S-PLOR-T

En el marco de la Tesis de Investigación de la Maestría de Ingeniería de Información (MINE) en la Universidad de los Andes, agradezco me colabores resolviendo esta encuesta de percepción sobre la herramienta S-PLOR-T que acabas de utilizar en la prueba de usuario realizada por MIGUEL ALFONSO FEIJOO GARCÍA.

SIGUIENTE

Página 1 de 5

Nunca envíes contraseñas a través de Formularios de Google.



EVALUACIÓN DE CALIDAD DE LA HERRAMIENTA: S-PLOR-T

*Obligatorio

Autorización

Autorizo al investigador MIGUEL ALFONSO FEIJÓO GARCÍA, estudiante de MAESTRÍA EN INGENIERÍA DE INFORMACIÓN (MINE) a obtener, tabular y analizar los datos proveídos en esta encuesta de percepción en el marco de investigación: TESIS, en la UNIVERSIDAD DE LOS ANDES – FACULTAD DE INGENIERÍA. *

- Sí Autorizo
- No Autorizo

ATRÁS SIGUIENTE

Página 2 de 5

Nunca envíes contraseñas a través de Formularios de Google.

EVALUACIÓN DE CALIDAD DE LA HERRAMIENTA: S-PLOR-T

*Obligatorio

Información General

Indica tu Nombres y Apellidos completos *

Tu respuesta

Indica tu Edad *

Tu respuesta

Indica tu Género *

- Mujer
- Hombre
- Prefiero no decirlo
- Otro: _____

Indica tu Rol actual: *

- Estudiante Pregrado
- Estudiante Maestría
- Estudiante Doctorado
- Profesor
- Otro

Indica el programa académico al que perteneces *

Tu respuesta

Indica la institución a la que perteneces *

Tu respuesta

En términos de "Análisis de Datos", ¿has tenido algún tipo de experiencia? *

- Sí
- No

En términos de "Análisis de Datos con características Espacio-Temporales", ¿te consideras un experto en el dominio? *

- Sí
- No

EVALUACIÓN DE CALIDAD DE LA HERRAMIENTA: S-PLOR-T

*Obligatorio

PREGUNTAS DE PERCEPCIÓN GENERAL

¿Qué tan interesante te pareció la herramienta S-PLOR-T? [Marca tu respuesta en una escala de 0 a 5, siendo 0 NADA INTERESANTE y 5 MUY INTERESANTE] *

0	1	2	3	4	5
○	○	○	○	○	○

¿Fue fácil el uso de la herramienta S-PLOR-T, en "términos gráficos"? [Marca tu respuesta en una escala de 0 a 5, siendo 0 MUY DIFÍCIL y 5 MUY FÁCIL] *

0	1	2	3	4	5
○	○	○	○	○	○

¿Fue fácil el uso de la herramienta S-PLOR-T, en términos de "análisis de los datos"? [Marca tu respuesta en una escala de 0 a 5, siendo 0 MUY DIFÍCIL y 5 MUY FÁCIL] *

0	1	2	3	4	5
○	○	○	○	○	○

¿Fue fácil ubicar el contenido y las funcionalidades de la herramienta? [Marca tu respuesta en una escala de 0 a 5, siendo 0 MUY DIFÍCIL y 5 MUY FÁCIL] *

0	1	2	3	4	5
○	○	○	○	○	○

A partir de los datos analizados, ¿el conjunto de herramientas de visualización y la relación entre ellas ayudó al análisis y hallazgos de resultados de interés? [Marca tu respuesta en una escala de 0 a 5, siendo 0 NADA y 5 MUCHO] *

0	1	2	3	4	5
○	○	○	○	○	○

¿Te pareció flexible la herramienta en "términos gráficos"? [Marca tu respuesta en una escala de 0 a 5, siendo 0 NADA FLEXIBLE y 5 MUY FLEXIBLE] *

0	1	2	3	4	5
○	○	○	○	○	○

¿Te pareció flexible la herramienta en "términos de analítica de datos"? [Marca tu respuesta con una X de 1 a 5, siendo 0 NADA FLEXIBLE y 5 MUY FLEXIBLE] *

0	1	2	3	4	5
○	○	○	○	○	○

A continuación, encontrarás un espacio para que indiques comentarios POSITIVOS o que te hayan parecido interesantes, cada uno de una frase. [Escribe tus comentarios en un listado en el siguiente espacio]

Tu respuesta

A continuación, encontrarás un espacio para que indiques comentarios que indiquen MEJORAS en la herramienta, cada uno de una frase. [Escribe tus comentarios en un listado en el siguiente espacio]

Tu respuesta

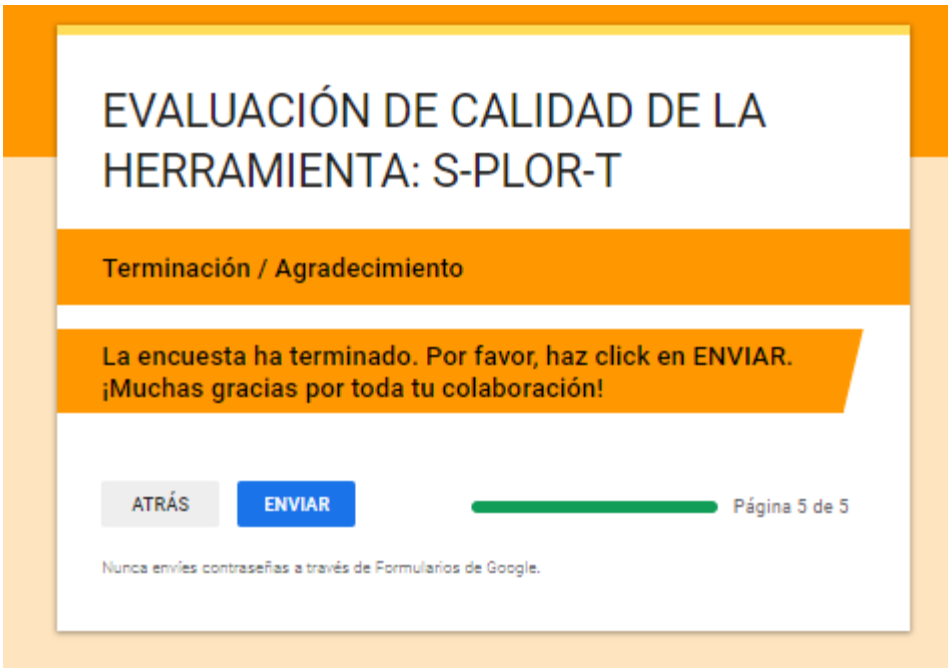
¿Consideras que S-PLOR-T es una buena herramienta para análisis exploratorio de datos con características espacio-temporales de forma flexible y agradable? *

○ Sí

○ No

Independiente de si tu respuesta fue SI o NO, indícanos brevemente por qué en el siguiente espacio:

Tu respuesta


EVALUACIÓN DE CALIDAD DE LA HERRAMIENTA: S-PLOR-T
Terminación / Agradecimiento
La encuesta ha terminado. Por favor, haz click en ENVIAR.
¡Muchas gracias por toda tu colaboración!
ATRÁS
ENVIAR
Página 5 de 5
Nunca envíes contraseñas a través de Formularios de Google.