Built Environment Profiles for Latin American Urban Settings

Fabian C. Peña*
Systems and Computing Engineering Department
Universidad de los Andes, Colombia

Andres F. Useche†
Industrial Engineering Department
Universidad de los Andes, Colombia

Abstract

Several characteristics of the Latin-American cities (LAC) make it especially suitable for studying the complex dynamics of the Built Environment (BE). The countries of the region are highly urbanized, with 61 to 93 percentage of the population living in urban areas, they have a large variation of inequality with a gini coefficient that varies from 0,34 to 0,63 and is the most violent region in the world, with 24.7 homicides per 100.000 inhabitants. These inequalities, urbanization and violence have allowed the development of some innovative and sustainable transport policies. To manage the complexity and diversity of LAC countries, it is necessary to develop profiles for cities (n = 371) and sub-cities (n = 1.436). It was performed a clustering model involving a wide set of attributes (n = 48) related to urban landscape, street design and transport, which are theorized or observed associated with public health. From the clustering model a set of profiles were identified, and expert users require a mechanism (interactive tool) to perform multidimensional analysis of profiles characterizing them in terms of the behaviour of the attributes used for learning the clustering model. This paper describes the proposal in terms of the design of this tool with focus on analysis of the data available, user tasks identified and visualization and interaction elements to fulfill each task.

Index Terms: Visual Analytics—Clustering Visualization—Urban Health—Built Environment.

1 Introduction

SALURBAL[1] is a 5-years project bringing together researchers from more than 10 countries to study how urban environments and policies impact publics health around Latin-American cities (LAC). For this purpose, 4 macro-objectives have been formulated. One of them is: "Quantify the contributions of the city's factors and urban neighborhoods to the difference in health levels inside each city and among different cities". To achieve this macro-objective, an analysis regarding to what Built Environment (BE) typologies emerge from multiple city attributes needs to be considered.

This paper covers the details of implementation and evaluation of a visual tool aimed to provide the capacity to user experts for analyzing the typologies previously determined by clustering algorithms from 2 groups of attributes: (1) urban landscape and (2) mobility infrastructure. Based on the visualization framework defined by Munzner [1], next sections give a deeper problem understanding and description of the target user profile, data available, user tasks to be considered, validation schemes, usability testing and conclusions. In addition, the main elements of the visual tool are provided in the context of some related work.

*e-mail: fc.pena@uniandes.edu.co
†e-mail: af.useche10@uniandes.edu.co

[1]https://drexel.edu/lac/salurbal/

2 Problem Statement

BE significantly affects the public health. For more than a decade, there has been an increase in the collaborations among public health, social sciences and urban planning researchers to study how BE impact health and health inequities. These fruitful multidisciplinary collaborations have provided visibility to many of the interrelated attributes of the BE that directly or indirectly affect health and health inequalities. Specifically, BE has been associated with mental health, non-communicable diseases, healthy behaviors, road traffic deaths and social inequities in health. Nevertheless, most of the evidence of BE and public health has been conducted in high income countries but few studies consider LAC.

One approach to managing the complexity and volume of the data describing the urban settings and environmental characteristics of LAC is to develop BE typologies for cities and sub-cities. These typologies are useful because they support the understanding of how multiple dimensions of the BE come together in these cities. They can also streamline the development of policy scenarios for future projections of urban development and growth.

For the BE group of the SALURBAL team is important to provide mechanisms to facilitate decision-making processes for user experts through an interactive visualization detailing the different typologies identified by clustering algorithms and their respective explanation in terms of the behaviour of the attributes considered.

3 User Description

The tool developed intend to serve as decision-making support mechanism in the contexts of town planning and public health. This fact does not imply that the tool cannot be used by an user without previous experience in these fields of knowledge. In this sense, two kind of users are defined: data analytics user and domain expert user. This classification have primary importance in Section 7 where the results of usability tests are discriminated and compared in terms of effectiveness of the tool to fulfill the tasks proposed.

A data analytics user is one with some experience and skills in topics such as information visualization, statistics, machine learning, among others, and may not have knowledge in the context of the data. Generally, a data analytics user could use this research as informative work to improve or develop more tools related to research in urban planning and urban health fields. A domain expert user is a researcher or practitioner working on fields such as urban planning or public health and may not have experience in data analytics. In general terms, the development of this tool pretend to have impact on policy-making contexts, where domain expert users have certain level of understanding of the problem and its implications, but don't have enough methodological approaches to achieve a better understanding and solution design of the problem.

4 Data Abstraction

Distinct aspects from BE and urban policies are considered. For this purpose, a set of socio-demographic, built environment, transport systems, among other attributes are available for LAC with more than 100.000 inhabitants. For each attribute category, three analysis levels can be performed, as described in Figure 1. The first or administrative level (L1Admin) corresponds to the political-administrative boundary defined by governmental entities. The L1Admin level is

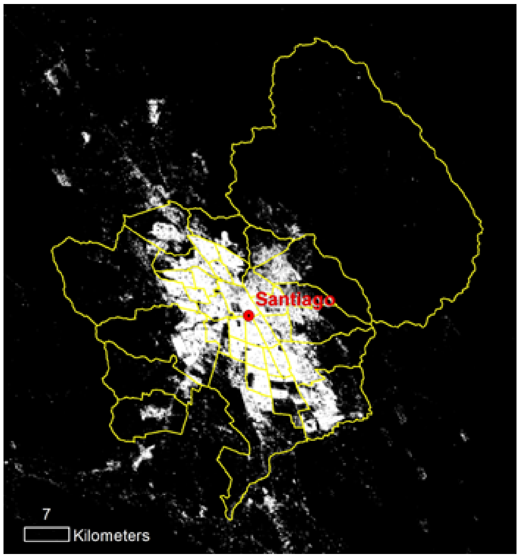

Figure 1: Example of analysis levels for Santiago de Chile city. The external yellow outline delimits the administrative area defined by governmental entities, named L1Admin level. Subdivisions of administrative area are the sub-cities, named L2 level. Finally, the white area is the zone of urban expansion and defines the unit actual boundaries, named L1UrbExt level.

the greatest analysis level and shows the dynamic of whole city. The second level (L2) corresponds to administrative subdivisions defined by each entity and represents the analysis level for the dynamics inside the city. The last level (L3) details neighborhoods as the most granular level of analysis. This third level is not finally considered because of availability of data for all units. The attributes are builded from satellite images and shape files. With this in mind, L1UrbExt analysis level is proposed consisting of the apparent extension of the urban boundary that is defined by zones where there are human settlements. The total amount of units for L1Admin and L1UrbExt is 371 cities and 1.432 sub-cities for L2, around 11 Latin-American countries. This information is provided by SALURBAL in **Table** format taken in an instant time (**Static**) where an **Item** corresponds to a geographical unit (city of sub-city previously defined), so additional derivation activities are not required.

For each analysis level, three **Domains** are specified to characterize the BE: urban landscape (12 attributes), street design (15 attributes) an transport systems (20 attributes). For urban landscape attributes, a patch as the minimum analysis unit is defined and corresponds to a grid of 100x100 pixels. This attributes measure the shape, fragmentation and isolation level by unit patches, in a normalized way. The street design attributes measure aspects related to density and shape of track network by unit. For both categories, the attribute type is **Quantitative Sequential**. Finally, transport system attributes indicate the presence of sustainable transport systems such as BRT, subway and bicycle in addition to phenomena as congestion and air quality. The presence of some of the sustainable transport systems is a **Categorical** attribute.

Position attributes (latitude and longitude) of type **Quantitative Diverging** for each unit are provided, so dataset type also can be considered as **Geometry**. In addition, we mention previously that BE typologies are identified by clustering algorithms by SALURBAL team. Typologies (or profiles) are **Categorical** attributes.

5 Task Abstraction

The tool developed must provide mechanisms to **Discover** multiple kind of insights from typologies previously identified by clustering algorithms, instead of just presenting information. The tasks described below have been refined after around 5 meetings with domain expert users who partially approve the mockup showed in Figure 2 and let construct the final visualization here. Because this is a functional mockup, some of the tasks have been validated and feedback in terms of complementary tasks to be included and visual design of the tool have been obtained and considered for mockup iterative adjustment. The tasks that user can achieve or perform with the visualization are:

- **Main task 1:** Characterize the profiles previously identified. This profiles must be explained in terms of attributes considered during modeling stage, so users need to **Compare** the **Distribution** of different attributes among profiles.
- **Main task 2:** For itself, the profiles are not meaningful at least that were related with some demographic or health outcome. Users need to **Compare** the **Distribution** of demographic or health outcomes among profiles.
- **Secondary task 1:** Users outside SALURBAL context need to identify how many cities or sub-cities from his country are included in the project. In other terms, users need to **Lookup Features** of a specific city or sub-city included in the analysis.
- **Secondary task 2:** The motivation to perform clustering methods to find profiles for Latin American cities and sub-cities is not clear for many experts. So it is important to show users the variability among cities and sub-cities in the specified domain. Users need to **Compare** the **Distribution** of different attributes for each country.
- **Secondary task 3:** Many cities have highest values for some attributes. Users need to **Identify** cities and sub-cities with **Extreme** values for some BE attributes.
- **Secondary task 4:** As an extension of Task 1, users need to **Compare** two specific attributes and validate the **Correlation** between them for each profile. Many attributes are highly correlated, so experts need to focus in less attributes to make their analysis.
- **Secondary task 5:** Select groups of units by a specific demographic or health outcome. Expert users can be interested in **Browse** or **Identify** cities or sub-cities with **Extreme** values for the outcome.
- **Secondary task 6:** Users need to see the identified cluster for his countries of interest and see the relation with specific demographic or health outcomes. For users, public policies could be developed based on specific results countries. Users need to Browse the cities or sub-cities of a specific country and **Compare** their **Similarity** in terms of demographic or health outcomes and the relation among the profile.

6 Proposed Solution

From data provided by SALURBAL team and tasks previously defined, we have developed a history telling visualization by using mainly marks of **Points** and **Areas**, and channels of both **Vertical** and **Horizontal Position**, **Color** and **Area Size**, as shown here. **Color Hue** is used in all interface elements to discriminate units belonging to each typology, complemented with **Use** encoding to show the distribution of units in the map by typology in a **Superimpose** facet (center-top). We follow the patterns of **Arrange** and **Separate**, **Order** and **Align** to characterize the distribution of the quantitative attributes discriminated by typology in the tickz chart.

To support the first secondary task, two dropdown elements are provided enabling users to lookup units by country or its own name.

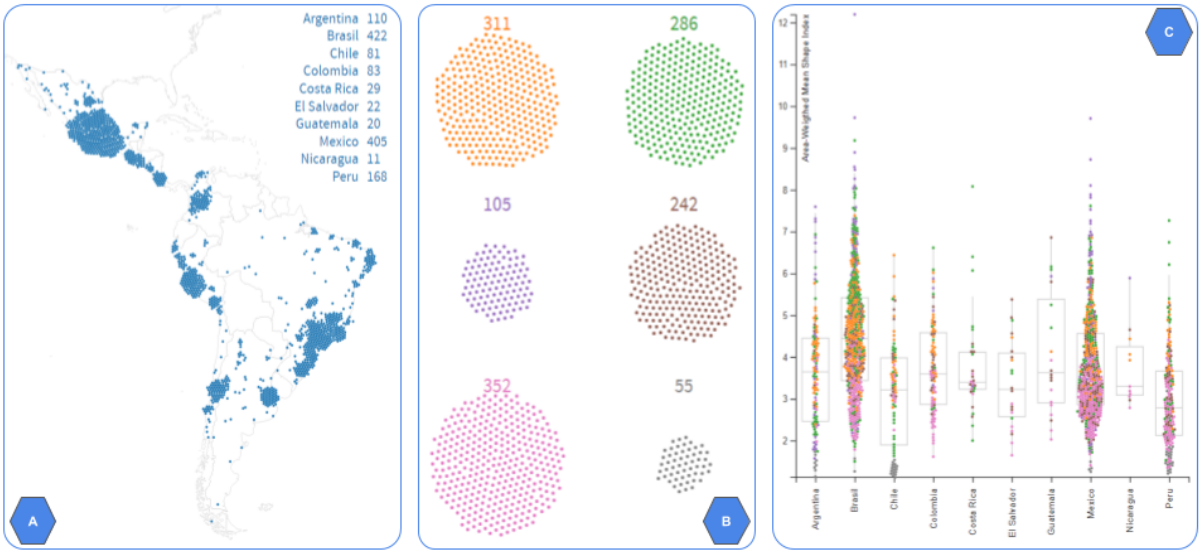

Figure 2: Final version of the interactive tool including some of the tasks described in Section 5.

All elements dispose of a cross **Filter** meaning that a change in one of them affects the data shown in the rest of the interface. For instance, user must be able to **Select** a range of values for certain attribute in the gantt chart and alter the treemap, the map and the summarizing table elements.

Main task 2 has not been included in the mockup because of data availability. Nevertheless, visual elements will consist of a scatter plot using **Point** as marks and **Position on common scale** as channels. Finally, complemented with channels for typology discrimination and interaction elements previously considered.

7 Usability Testing

Initial usability testing has been performed in two stages: the first one is more informal and is focused in getting direct feedback from domain expert users involved in SALURBAL project. This stage is developed during SALURBAL meeting in Mexico. The second stage comprises a usability testing based on initial exploration of the tool, specific tasks to be evaluated and final structured survey response by users.

7.1 Stage 1: Mexico Meeting Interviews

The purpose of the interviews in SALURBAL meeting was to collect feedback about the profiles research. The main objective with the feedback was to ensure the interpretability of the profiles with the visualization. To get meaningful results from the interviews we implemented the following three steps procedure: **1)** Present the profiles motivation, methods and results, **2)** Present the visualization and let the user interact with it and **3)** Collect the users' comments on the research and visualization. For practical terms, we are going to focus only in third step results. To achieve this step we let users to interact with the visualization and in printed versions the users could make notes and draws to highlight or delete aspects they wanted.

In general, the users focused or highlight three aspects of the visualization: **a)** color range, **b)** comparison of units by indicator, and **c)** use of measurements. In the aspect **a**, the users argued that the colors used are not properly to visualize the results of the profiles. For them, colors like red or green had a intrinsically meaning that could generate incorrect associations (e.g red is bad, green is god). In aspect **b**, the users mentioned that the form of the indicator distribution does not allow to easily compare it. They think that if a reference point of comparison does not exist, the comparison process is more difficult. The suggestion for this problem is to provide two mandatory and one variable changes. The first mandatory change is to include the unit of measurement for each indicator. The last mandatory change is to include the meaning of the extreme values for each indicator (e.g a high level of circuity average indicate a low directness of the streets). The variable changes imply to include some statistical measurement to make comparison in and between profiles (e.g include in the visualization the mean, median or mode). In the aspect **c**, the users suggest that descriptive statistics that change in terms of the comparability level could facilitate the comparison process. These levels make comparison between: units in different profiles, units in the same profile, units in the same country and a selection of units.

7.2 Stage 2: Structured Usability Testing

This study was performed in two stages: the first stage corresponds to make test with non-experts or analytics users. The second stage corresponds to make test with field experts or analytics users. The usability test is structured by: **(1)** an initial exploration of the tool and subsequent explanation about the context of this, **(2)** the assignment of five tasks as questions to be performed by users and **(3)** a final survey filling. The survey includes user demographic aspects, task fulfillment in terms of easiness, usefulness and satisfaction, and general problems founded and recommendations given by users. Details about the testing script can be founded here, and demographic statistics in the Appendix A.

For the first stage, the survey include the following task to evaluate the usability of the visualization:

- **Q1:** Select the typology 5 and say what attribute is in a different range of values with respect to other typologies.
- **Q2:** Find the units belonging to Nicaragua and say how their attributes are behaved with respect to the total of the units.
- **Q3:** Select a group of units belonging to north of Mexico, preferably few, and compare them in terms of belonged to typologies and attribute values.

Usability testing results for first stage in each question are showed in Figure 3. For this little group of users, the tool seems be enough to fulfill the tasks of interest.

Some of problems and recommendations from this second group pf users can be summarized as follow:

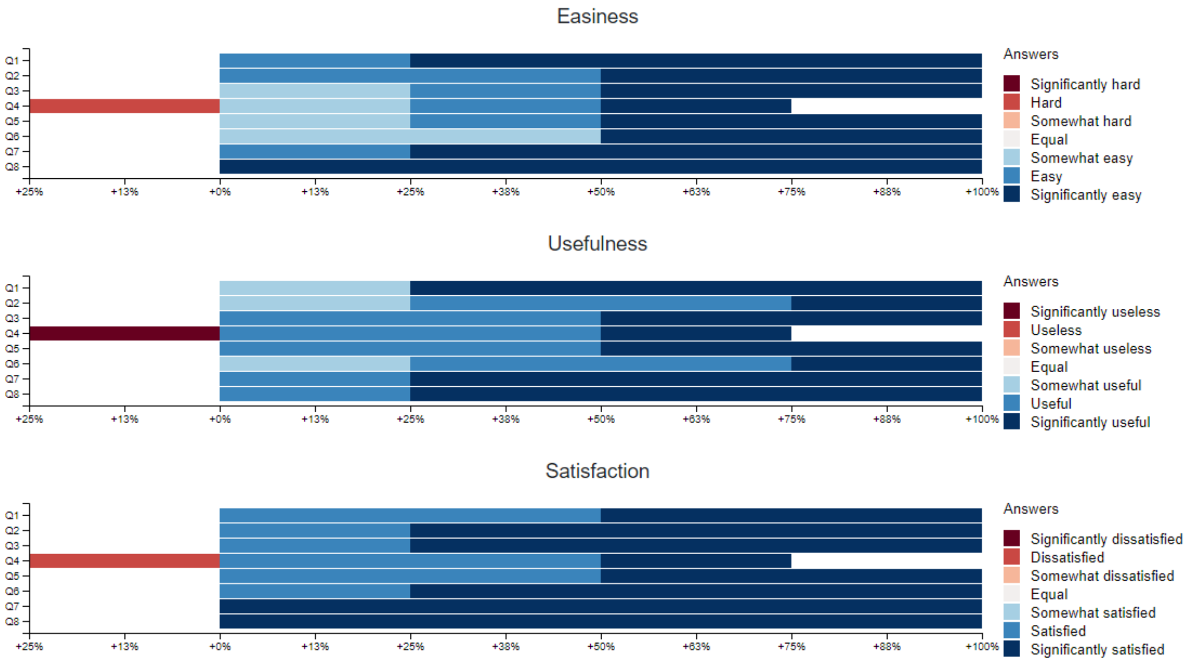

Figure 3: Likert scale for the three questions considered.

- Attributes are showed without units, making difficult put in context the analysis performed by the user.
- Additional statistics in terms of units currently selected by the user can be useful to improve the analysis of specific groups of units.
- Too much information to show. Units with similar attributes or geographic distribution are overlapped. Mechanism to support the readability of near units must be implemented.

For the second stage the survey was modified to include all tasks defined in section 5 with these specific tasks:

- **Q1:** Highlight the units belonging to Colombia.
- **Q2:** What city has the highest number of patches?
- **Q3:** Which country presents the highest variability for the area-weighted mean shape index variable?
- **Q4:** Which are the two profiles with the highest number of cities from Colombia?
- **Q5:** Identify the profile with the highest patch density distribution.
- **Q6:** Identify which couple of different variables present the highest positive correlation in profile two.
- **Q7:** Identify what is the profile with the bigger cities.
- **Q8:** Identify what is the city with the highest population density and in which profile is located?

From the final interviews and with the survey's results we can say the following about the easiness, satisfaction and usefulness of the proposed visualization:

- Two of the four interviewed users thought that visualization allowed them to easily solve the proposed tasks.The task four was hard to solve for only one of the four users.
- Two of the four interviewed users thought that visualization was useful to solve the proposed tasks. Only one of the four users thought that the visualization was significantly useless to solve the task four.
- Three of the four interviewed users were satisfied with the way the visualization allowed them to solve the proposed tasks. Only one of the four users was dissatisfied with the way the visualization allow to solve the task four.

In general terms, for the interviewed users the visualization allowed them to solve the proposed tasks. Another important aspect is that visualization highly achieved the goal of present the research work and told a history to users. The most difficult aspect of this research study is to achieve that non-expert user could understand the motivation about developing typologies for understanding built environment. The interviewed users think that the visualization can help to improve the understanding of this challenge imply.

8 Conclusion

In academic contexts, present the result of a research study can represent a challenge for the researchers. The project of developing Built Environment Typologies for Latin American Urban Settings is some of these challenging cases, where the methodological approaches are highly complex to understand and the results for itself are not completely meaningful.

The Tamara Munzner framework is a useful language to traduce the requirements of the research study to specific tasks that can be transformed into visualizations. This visualization allows the final users to get insights about the results of the study and generate policies, in this specific study, to improve health outcomes in Latin American countries.

References

[1] T. Munzner. *Visualization Analysis and Design*. 2014.

Appendix A Demographic Analysis

This appendix is focused on detail the demography of each user who participate in the usability testing. For this study, just attributes such as genre, age and education level are considered and summarized in Figure 4. All users participating in the study are domain expert users.

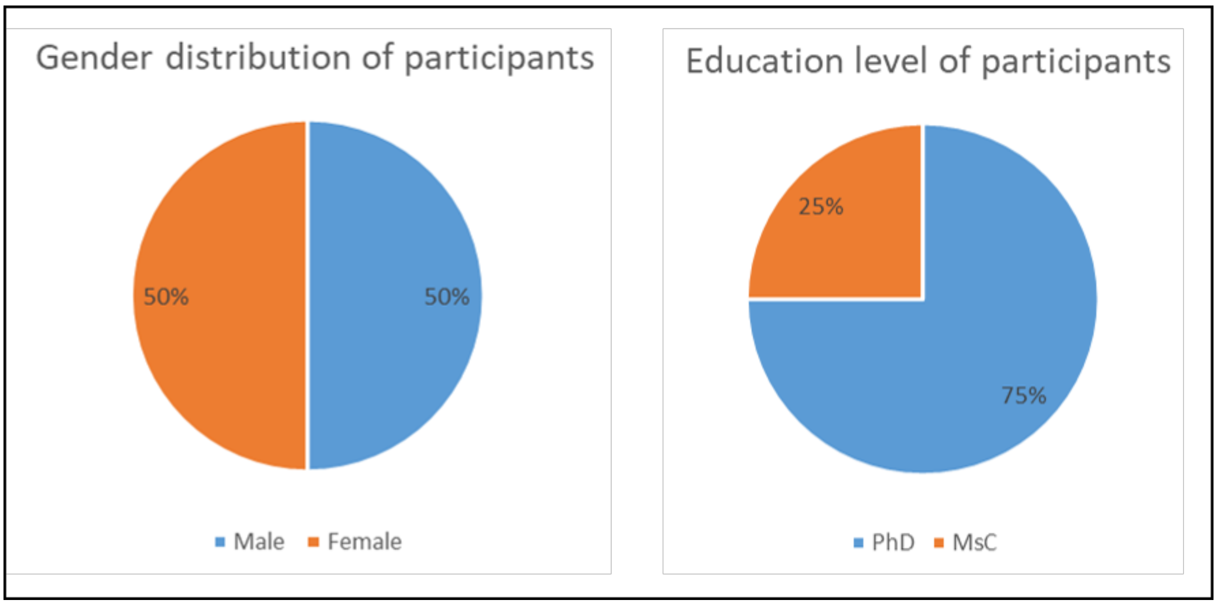

Figure 4: Genre and education level of users participating in the study.