# Remote Usability Assessment of Topic Visualization Interfaces with Public Participation Data: A Case Study

## Ivania Yovanovic*, Julian Goñi**, Constanza Miranda***

* Engineering School, Pontificia Universidad Católica de Chile, iyovanovic@uc.cl, Author ORCID Nr: 0000-0001-8001-5939
** DILAB Engineering Design, Engineering School, Pontificia Universidad Católica de Chile, jvgoni@uc.cl, Author ORCID Nr: 0000-0002-0170-3416
*** Biomedical Engineering Department, Johns Hopkins University, constanzamiranda@jhu.edu, Author ORCID Nr: 0000-0002-9110-2832

Abstract: Citizen participation often faces challenges of transparency and accountability. Visualizations's usability becomes key for public consultation activities. The tree map is frequently used to disseminate data and to give it back to the population. The purpose of this study is to understand how tree maps and stacked barcharts differ in terms of effectiveness, efficiency, and satisfaction in tasks, like solving topic categorization and comparison analysis tasks. An experimental design was used to examine user performance based on a task-based usability test. 34 participants interacted remotely with data visualizations from an open 2016 participatory constitution-making process. The ANOVA showed that stacked barcharts work significantly better for comparison tasks than the tree map, but there are no significant differences in regards to categorization tasks. Public participation initiatives should first determine what cognitive operations their users perform before deciding which visualization interfaces will be more useful for the intended public.

Keywords: Topic visualization, usability, human-computer interaction, public consultation, citizen participation

## 1. Introduction

'Citizen participation' is a broad category that describes citizen involvement in public decision making (Irvin & Stansbury, 2004). Depending on the conceptual framework you sit upon, 'citizens' may either be individuals or organized communities, and 'participation' may involve either observation or direct power (Baum, 2015). Additionally, the scope of participation can be at the community or national level and in various spheres of government and society (Cardullo & Kitchin, 2019). One of the earliest definitions of citizen engagement was proposed by Arnstein (1969). The author

formulated a conceptual ladder with eight involvement levels. These range from manipulation as its weakest form; to the most complete which is direct citizen control. Nowadays, there is a tendency among political science scholars and practitioners to demand for more active, direct and substantial citizen engagement in policy-making (Dryzek, 2002; Floridia, 2018).

Public participation in policy-making ranges from local government budgeting to constitution-making (Bobbio, 2019). Citizen participation in constitution-making processes is a growing tendency (Pacheco i Canals, 2019). Over 70 percent of new constitutions promulgated in the last decades have involved citizens (Maboudi & Nadi, 2016). The forms of citizen engagement, ranges from the election of constitution drafters to public consultations over the content of constitutions (Negretto, 2017). Despite the fact that public participation in constitution-making is highly desirable for achieving legitimacy, there are persistent challenges in the process for managing institutions, especially relating to inclusivity, transparency and accountability (Gluck & Brandt, 2015). Some countries have started to incorporate new digital technologies to facilitate that process (Gluck & Ballou, 2014).

Beyond the topic of constitution-making, several countries have used Artificial Intelligence (AI) techniques to organize and classify the content of their public consultations (Noveck, 2018). Topic modeling is a machine learning technique that has been useful for this purpose (Hsiao, Lin, Tang, Narayanan & Sarahe, 2018). This technique identifies and organizes the underlying themes in opinions automatically (Blei, Ng & Jordan, 2003). However, the end results of this sort of automated analysis may be difficult for citizens to understand (Cai, Sun & Sha, 2018). This is why many public participation exercises accompany data analysis with visualizations tools to increase the usability of the information gathered (Roque de Oliveira & Partidário, 2020).

Different visualization interfaces have been developed to represent the topic models (Karpovich, Smirnov, Teslya & Grigorev, 2017). The aim of those interfaces is to allow the user to analyze topic models. The tree map is one of the most used visualizations to present an overview of topics organization and allow categorical and quantitative analysis (Sukhija et al., 2019); how many topics, what are more important, and what are the categories between them. The biggest problem with tree maps is that using size to encode the data makes it difficult to make precise quantitative comparisons versus using length of a bar (Li et al., 2019). In that sense, other visualization strategies such as stacked bar charts, may offer a better way of encoding data for precise comparisons.

This article explores and compares usability characteristics of two types of topic visualization interfaces using data from a public participation exercise. We compare the usability of the stacked bar chart and the tree map visualization techniques within an interface complemented by two maps: a choropleth map and a point density map, since the use of maps has previous evidence of its validity in supporting georeferenced information in analysis tasks (Jankowski & Nyerges, 2001). For the topic model, the Citizen Consultation of the 2016 Constituent Process in Chile was utilized as a case study. This particular consultation brought together more than 200,000 citizens who gave their opinion for the content of a possible new constitution (OECD, 2017). This inquiry is based on the interest in the design of interfaces in the field of human-computer interaction (HCI) as described by researchers such as Polasanapalli and Buggareddy (2020).

## 2.   Theoretical Framework

## 2.1.   Topic modeling and topic visualization

With the rise of computer science, text analysis tasks have become automated (Humphreys & Wang, 2018). Artificial intelligence (AI) is the scientific field for it. With a large amount of training data, AI allows the understanding of different perspectives that underlie documents (Radford, Narasimhan, Salimans & Sutskever, 2018). One of the most common and useful perspectives for understanding documents is through their main topics. The process of learning and extracting these topics is called topic modeling (Brynjolfsson & McAfee, 2017). This machine learning technique applies probabilistic models to extract underlying themes from large collections of documents. (Blei, Ng & Jordan, 2003). In topic modeling research, a topic is defined as a set of words that has a high probability of appearing together (Jelodar et al., 2019).

Latent Dirichlet Assignment (LDA) is one of the most popular methods in topic modeling (Sukhija et al., 2019). LDA has been used for opinion modeling from social media (Parra et al., 2016; Sun, Luo & Chen, 2017; Xu, Liu, Xuan, Chen & Mei, 2017), customer reviews (Calheiros, Moro & Rita, 2017) and public consultations (Hagen, Uzuner, Kotfila, Harrison & Lamanna, 2015). Although it has been useful for experts in the area of computer science, the results of the model are not very intuitive for ordinary users (Cai, Sun & Sha, 2018). That's why in recent years, several topic visualization interfaces have been developed to simplify and communicate what people opine in social media, about products and services or in public affairs. (Karpovich, Smirnov, Teslya & Grigorev, 2017). One of the main objectives of this type of visualization is to provide users the options to explore the set of documents and identify what topics they relate to (Ganesan, Brantley, Pan & Chen, 2015).

According to Yang & Qu (2017) there are three guidelines for building efficient and user-centered  topic visualization interfaces (i) An overview of the organization of the topics that allows the user to understand how many topics there are, what is more important and what are the categories, (ii) a geographic perspective view of the topics that allows the user to explore geolocated data; (iii) a particular view of opinions by topic that allows the user to review the compilation of opinions or an extract from them.

Different methodologies have emerged to follow these guidelines and create user-centered visualizations. One of them is the one proposed by Tamara Munzner (2009). This method combines four areas of knowledge: anthropology/ethnography, design, computer science, and cognitive psychology. Munzner proposes designing visualizations through three fundamental questions that are asked iteratively.  These questions are:
1)   What information will be displayed on the visualization?
2)   Why does the user need the information?
3)   How will the information be encoded?

To design topic visualization interfaces, questions 1 and 2 can be answered through the guidelines previously outlined. For question 3, Mackinlay (1986) suggests coding the attributes with the best-ranked channels according to the principle of effectiveness.

## 2.2.   Visual encoding for topic visualizations

Data representation is a main component of information visualizations. The fundamental focus of data representation is mapping from data values to graphical representations (Cleveland & McGill, 1985). Visualization designers use elementary graphical units called visual encodings to translate data to visual elements (Kim & Heer, 2018). There are different visualization alternatives to encode the same information. It is important that these alternatives meet expressiveness criteria (Mackinlay, 1986). Given two graphical languages that express some information, the obvious question is which language involves a design that specifies the more effective presentation. Unlike expressiveness, which only depends on the syntax and semantics of the graphical language, effectiveness also depends on the capabilities of the users's personal characteristics (Qu & Hullman, 2016), such as cultural background.

Among the most popular visualizations to present an overview of topics distribution and categorization is the tree map (Long, Hui, Fook & Zainon, 2017). In it, the importance of each topic over the total quantity is encoded by the area of each square. The larger the area of the square, the greater the importance of the topic in question (Vernier, Telea & Comba, 2018). An alternative for this type of data is stacked barcharts, where the importance of each topic over the total quantity is encoded through the position and the length of the bar (Howorko, Boedianto & Daniel, 2018). In both cases, tree map or stacked barchart, the categories of the topics are encoded through color tones.

Several researchers have examined empirical user performance with visual encodings (Ware, 2019), validating and refining effectiveness rankings originally proposed by Bertin (1983). Cleveland & McGill (1985) conducted human-subjects experiments measuring decoding error across encoding channels, for example finding that position encodings outperform length encodings when comparing proportions. In addition to studies of univariate encodings, researchers have examined interactions between visual encoding channels (Garner & Felfoldy, 1970; Ware, 2019). For example, integral visual channels (e.g., color and size) may facilitate decoding when used to redundantly encode the same field (Demiralp, Bernstein & Heer, 2014).

A more recent trend is the study of tasks involving the perception of visual information. A large body of earlier work (Correll & Gleicher, 2014; Szafir, Haroz, Gleicher & Franconeri, 2016; Dimara, Bezerianos & Dragicevic, 2017) has studied the effectiveness of visualization types with common design configurations for a selected number of tasks. These studies usually focus on specified users, the specific ones for whom the visualization is designed, and on the specified context of use and the precise environment in which the users will use it (Barnum, 2020). Usability evaluation is the method that has been used to measure the encompass of tasks, users and context (Kim & Heer, 2018).

## 2.3.   Usability as an evaluation method of topic visualization interfaces

Within the field of human-computer interaction, usability has become a highly used concept when evaluating visualizations (Umar, Bakhat and Hassan, 2020). Usability has become part of the set of standards established by the International Organization for Standardization (ISO). Usability is understood as the degree to which a product can be used by specific users to achieve goals with

effectiveness, efficiency and perceived satisfaction in a specific context of use (ISO 9241). Therefore usability is understood and measured in three dimensions:

1) Effectiveness refers to the performance of tasks, to the precision with which the user meets their objectives. It is usually evaluated based on precision such as error rate or success rate (Howorko et. Al, 2018).

2) Efficiency refers to the relationship between the level of effectiveness and the consumption of resources. For example, metrics of the number of clicks and time to complete the task are used (Zia, Guo, Zhou, Essa & Jarc, 2019).

3) Perceived satisfaction refers to the comfort reported by the user after solving a task. Satisfaction level metrics are used on Likert scales (Isaac, Abdullah, Ramayah & Mutahar, 2017).

Usability evaluation methods are used to evaluate the interaction of the human with the computer for the purpose of identifying aspects of this interaction that can be improved to increase usability (Lewis, 2019). Through usability evaluations, the most appropriate elements for an interface are identified. In them, the degree of effectiveness, efficiency and satisfaction with which users can use the system designed to solve certain tasks is measured (Manresa-Yee, Amengual, & Ponsa Asensio, 2014).

The main goal of topic model data analysis is to identify topic clusters, explore patterns and locate representative opinions of these topics. The topic model analysis process can be viewed as a set of tasks and operations needed to meet the goals of the data exploration. These tasks involve a number of more specific activities and operations that users will perform such as (Karpovich, Smirnov, Teslya & Grigorev, 2017):

1) Identification of how topics are categorized.
2) Identification of which category a certain opinion belongs to.
3) Comparison of the relative importance between topics.
4) Comparison of relative importance of a certain topic at different spatial locations.

To complete the tasks described above, the user will have to execute a number of visualization operations during the exploration process. Several authors have suggested taxonomies for visualization operations (Qian, Wachowicz, Peuquet & MacEachren, 1997; Zhou & Feiner, 1998; Ogao & Kraak, 2002). The most comprehensive list (Keller & Keller, 1992; Wehrend & Lewis, 2000) includes identify, locate, categorize and compare:

1) Identify: to establish the collective characteristics by which an object is distinctly recognizable.
2) Locate: to determine the absolute or relative position.
3) Categorize: to place in specifically defined divisions in a classification; this can be done by colour, position, or type of object (shape).
4) Compare: to examine so as to notice similarities, differences, order.

In conclusion, several advances have been made in modeling and visualizing public opinion. However, recent studies tend to focus mostly on social media opinion analysis (Steinskog, Therkelsen & and Gambäck, 2017; Yu, Xu, Wang & Ni, 2019; Kant, Weisser & Säfken, 2020). The modeling and visualization of opinions from institutional citizen participation' instances is still missing more attention (Mukherjee, 2019). That way, it would become more manageable to address the growing

interest and demand towards a more active, direct and substantial public participation in policy-making (Dryzek, 2002; Floridia, 2018).

On the other hand, usability evaluation encompasses three dimensions: effectiveness, efficiency and satisfaction. But since the beginning of visualization evaluation, mostly the effectiveness has been measured (Szafir, Haroz, Gleicher & Franconeri, 2016). Moreover, despite the fact that satisfaction was important when the ISO standard was developed, it has just recently begun to be considered (Barnum, 2020). Furthermore, about half of the studies in the area do not include evaluations of any kind (Elmqvist & Yi, 2015). Usability studies are needed to evaluate and validate the relevance of visualization interfaces (Yang et al., 2017).

In sum, it is not enough to use visualization strategies. Rather, the effectiveness of public processes requires that institutions provide evidence of the relevance of these visualization tools. Our study explores usability measures in visualization interfaces for topic modeling of citizen consultations in a case of participatory constitution-making in Chile. To achieve this objective, our research question is: what are the differences in effectiveness, efficiency, and perceived satisfaction between interfaces including tree map and stacked barchart in HCI challenges?

## 3. Materials and Methods

### 3.1. Case study: Topic model of Chilean public consultation

In 2016, the Government of Chile convened a citizen consultation to discuss the contents of a possible new constitution. More than 200,000 citizens gave their opinion about its content (OECD, 2017). The methodology developed by the government established four lines of debate: (1) values and principles, (2) rights, (3) duties and responsibilities, and (4) State institutions (Raveau, Couyoumdjian, Fuentes-Bravo, Rodriguez-Sickert and Candia, 2020).

A guiding question was defined to address each line. For each question, a short argument had to be written to justify the answer. The content of the consultation was processed by the Systematization Committee, an autonomous body external to the Government. Its methodology consisted of a manual work of syntactic unification of each opinion, without modifying its semantics (Fierro, Fuentes, Pérez and Quezada, 2017). The results were consolidated in an open database called Bases Ciudadanas (Becker Castellaro, 2018).

The size of Bases Ciudadanas was reduced by 9.8% after cleaning empty cells. Python, NLTK Toolkit and Gensim libraries were used for preprocessing. Capitalization, accents, and punctuation have been removed through tokenization. Stop words, words such as articles, pronouns and prepositions that do not add value to the meaning of the text were removed. The words in the third person changed to the first person and the verbs in the past and future changed to the present through a stemming process.

After preprocessing, a dictionary was built with all the words present in the opinion attribute of Bases Ciudadanas in order to train the model. In that way the model could 'learn' the words and their relation with others. The number of topics (k) is an important parameter to run the model,

which it has to be set beforehand. In order to set the optimal number of topics, many LDA models with different values of 'k' were run. Finally, the k's election was based on the highest coherence value (c_v). The optimal number of topics were 15 and the average training time was 26 minutes.

## 3.2. Topic visualization interfaces

Figure 1 shows the user interfaces for interacting with topics and opinions. Each version of the interface system has three focus areas. The upper left part is a topic distribution visualization that allows identifying the categorization of topics and the relative importance between them. The lower left part is a topic-region visualization that represents the spatial distribution of each topic. The right part is a topic-opinion visualization that allows to explore and interact with opinions in detail.

For the visualization of topic distribution, the 15 topics were classified into 5 categories, where each category is represented by a different color tone. When selecting a topic, the other two visualizations are filtered according to the topic in question. The topic-region visualization changes the color saturation of each region to represent the relative importance of that topic. Also, when selecting a region, the other two visualizations are filtered based on that region. In the topic-opinion visualization each point represents an opinion and these are coded according to the color of its most representative topic. When hovering over a point on the map, a pop-up window displays the opinion, its commune and region.

*Figure 1: User interface for the exploration of the model of topics on the opinions of the citizen consultation of the 2016 constituent process in Chile. (a) Interface A: tree map + choropleth map + point density map; (b) Interface B: stacked barchart + choropleth map + point density map*



(a)                                              (b)
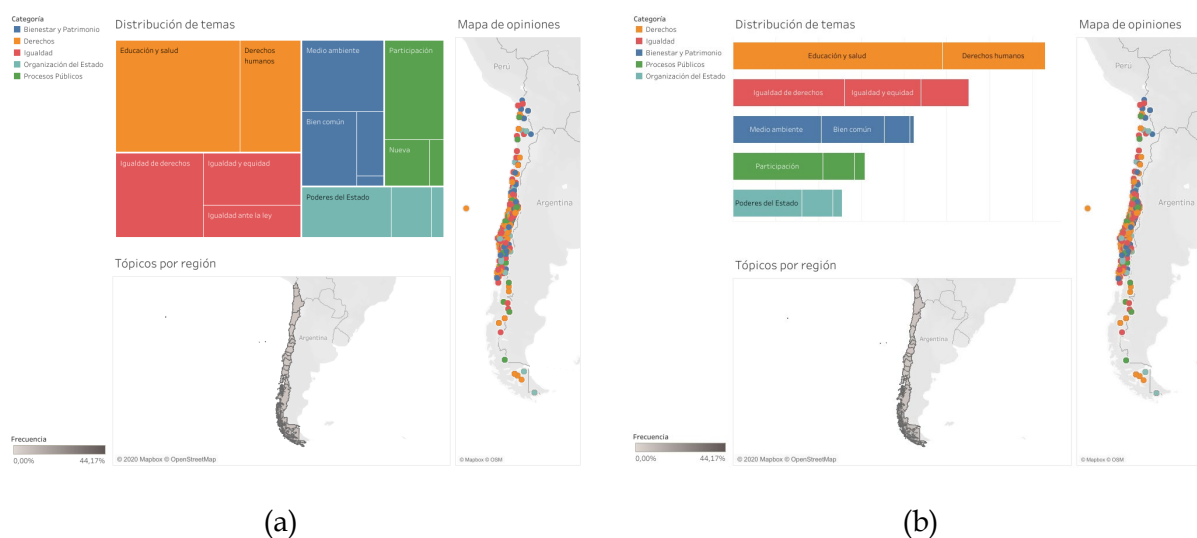
## 3.3. Instrument

For the data collection process, a Task-Based Usability Test was used (Corry, Frick & Hansen, 1997). This kind of test is designed to explore users' performance and perceived satisfaction given specific

tasks and interfaces such as categorize and compare. Setting up a usability test involves carefully creating a scenario, wherein the person performs a list of tasks using the interfaces being tested. Due to the COVID-19 health crisis, a remote usability testing was conducted (Barnum, 2020). This usability testing works well with any of the popular online meeting tools. In this case, the remote usability test was done through Zoom.

The test is structured in 4 sections covering: (1) topic categorization, (2) opinion categorization, (3) overall comparison and (4) comparison by region. The topic categorization section includes a question of alternatives: "Which topic is in the same category as "human rights"?. To answer this question two alternatives were provided. This section is focused on the use of the tree map or stacked barchart visualization.

The opinion categorization section includes a question of alternatives: "What topic could the following opinion belong to? "Creation of regional laws and regulations should enable the administration of own resources in a decentralized manner"?. To answer this question two alternatives were provided. This section is focused on the complementary use between tree map or stacked barchart, and point density map.

The overall comparison section includes a question of alternatives: "What is the most frequent topic between "New Constitution" and "Cultural Heritage"?". To answer this question two alternatives were provided. This section is focused on the use of the tree map or stacked barchart visualization.

The comparison by region section includes a question of alternatives: "In Los Lagos region, what is the most frequent topic?". To answer this question two alternatives were provided. This section is focused on the complementary use between tree map or stackedbar chart, and choropleth map.

After each section, participants had to grade 3 statements of the After Scenario Questionnaire (ASQ) (Lewis, 1995). Grading was in terms of how satisfactory it was using the interface to solve the task on a seven-point Likert scale (from 1 = "totally disagree" to 7 = "totally agree"). The statements were:

1) In general, I am satisfied with the ease of completing tasks on this interface.
2) In general, I am satisfied with the amount of time it took to complete the tasks on this interface.
3) In general, I am satisfied with the information provided to complete the tasks.

### 3.4. Participants

The participants in this study were a group of 34 engineers recently graduated from an university in Chile. To determine the number of participants GPower Software was used. Through a Power Analysis (Mayr, Erdfelder, Buchner & Faul, 2007), the sample size was determined assuming a significance level ($\alpha$) of .05, a statistical power (1-$\beta$) of .8 and an effect size *(d)* of .25.

The age of the participants was between 25 and 30 years old. It was limited to this age range due to its advantage in digital skills and to avoid effects associated with age (Hanson, 2011). The average age of the participants was 27 years old. Regarding the gender of the participants, 19 (56%) were

women and 15 (44%) were men. It was considered as an exclusion criterion having a handling or experience with data visualizations greater than 5 on a scale of 1 to 10. All the participants used the same equipment and carried out the assigned tasks in a time of 30 minutes.

None of the participants received financial compensation. Participants were asked to sign a consent form and were informed that they could leave the study at any time.

## 3.5. Experimental design

A design of 2 types of interface and 4 types of tasks, intra-subjects design was used. Each participant used both interfaces to perform categorization and comparison tasks through the Task-Based Usability Test. The participants were first briefed about the interfaces and were clarified on the tasks that they had to perform.

Two versions were performed to counteract cognitive fatigue and to control associated learning effects (DePuy & Berger, 2014). Table 1 shows the 2x2 Latin Square used to determine the order of interface types used by participants in the study. Participants assigned to version 1 first used Interface A and then Interface B. Participants assigned to version 2 first used Interface B and then Interface A. In total, 8 questions had to be answered per participant, 4 with interface A and 4 with interface B.

*Table 1: 2x2 Latin square for interface order by version*

| Version | Interface 1 | Interface 2 |
|---------|-------------|-------------|
| 1 | A: treemap + choropleth map + point density map | B: stacked barchart + choropleth map + point density map |
| 2 | B: stacked barchart + choropleth map + point density map | A: treemap + choropleth map + point density map |

The metrics considered to perform the usability evaluation in this study include performance measures (task success rate, task completion time), as well as self-reported measures (perceived satisfaction) (Sørum, Andersen & Vatrapu, 2012). The success rate was measured through the number of correct answers over the total of answers per task and per interface. Completion time was measured through the time (in seconds) taken by participants to complete a specific task using a particular interface. And finally, the perceived satisfaction was measured through the comfort level reported by the participants per task and per interface. The success rate and completion time were recorded automatically each time a participant responded to a question. The perceived satisfaction was collected through the ASQ.

### 3.6.   Data Analysis Tools

The study' aim was to explore the effects of two factors (type of interfaces and type of tasks) on success rate, completion time, and perceived satisfaction. The comparison between the two interfaces (A: Treemap + Choropleth map + Point density map versus Stacked barchart + Choropleth map + Point density map) was performed through a graphical analysis and statistical hypothesis tests.

To visually compare the results, the means and standard deviations of the performance measures across two task types and two different interface types were plotted on a bar chart. Although this type of comparison is not conclusive, it may help to identify trends on performance measures and (potential) differences between interfaces.

Given the setup of the experiment, the mean differences between the two factors were measured using two-way repeated measures analysis of variance. There were no significant outliers in any cells in the design. The dependent variable complies with the assumption of normality according to the Shapiro-Wilk test. Mauchly's test showed that the data met the assumption of sphericity. When the values were significant, post hoc comparisons with Bonferroni adjustment were made. The experimental data were analyzed with R Studio.

## 4.  Results

The main objective of this study was to compare usability metrics patterns in categorization and comparison tasks by two interfaces (A: including a tree map versus B: including a stacked barchart).

Figure 2 shows the mean for success rate, completion time and perceived satisfaction for each task of the Usability Test by interface. The behavior of the usability metrics follows a similar pattern in both interfaces, in which the mean were relatively close in categorization tasks (T1 and T2), and relatively far in comparison tasks (T3 and T4). This means that there were not major differences  in the success rate (see Figure 2a), completion time (see Figure 2b) and perceived satisfaction (see Figure 2c) between interfaces in categorization tasks, but there were differences between both interfaces in comparison tasks.

*Figure 2: Mean for (a) success rate, (b) completion time and (c) perceived satisfaction with half a standard deviation by interface (A: interface including a tree map versus B: interface including a stacked barchart) for each task (T1, T2: categorization tasks and T3, T4: comparison tasks)*

(a)                                                                        (b)
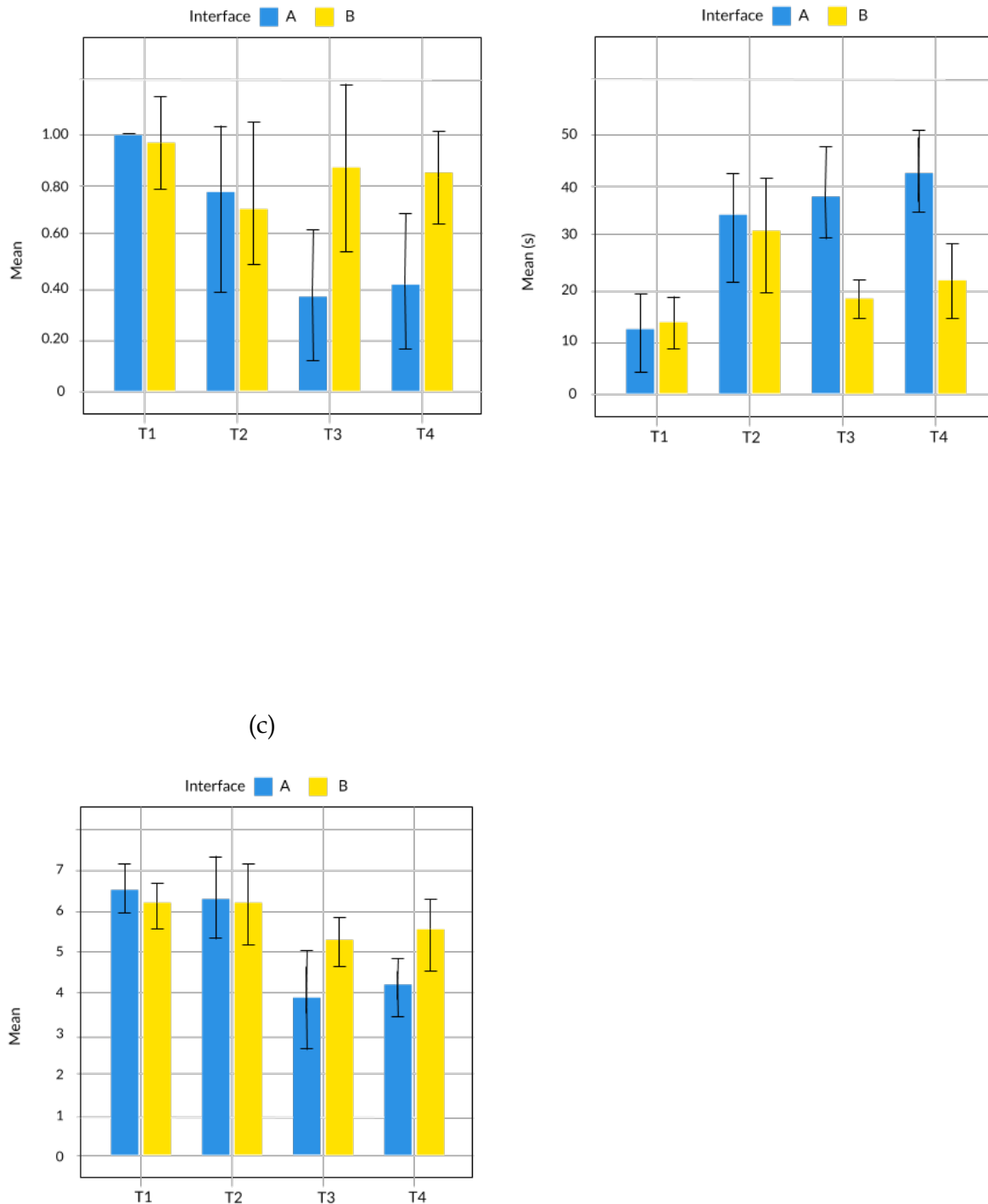
(c)



Figure 3 shows the means for success rate, completion time and perceived satisfaction for each task of the Usability Test by interface on an overlapping line chart. There is a common tendency where lines do not appear to be parallel and in some points they are even crossing. This could suggest an interaction between the factors where, success rate (see Figure 3a), completion time (see Figure 3b), and perceived satisfaction (see Figure 3c), have a relative behavior according to the task being performed and the interface being used.

*Figure 3: Line graph of the mean for (a) success rate, (b) completion time and (c) perceived satisfaction by interface (A: interface including a tree map versus B: interface including a stacked barchart) for each task (T1, T2: categorization tasks and T3, T4: comparison tasks)*

(a)                                                                                      (b)



(c)
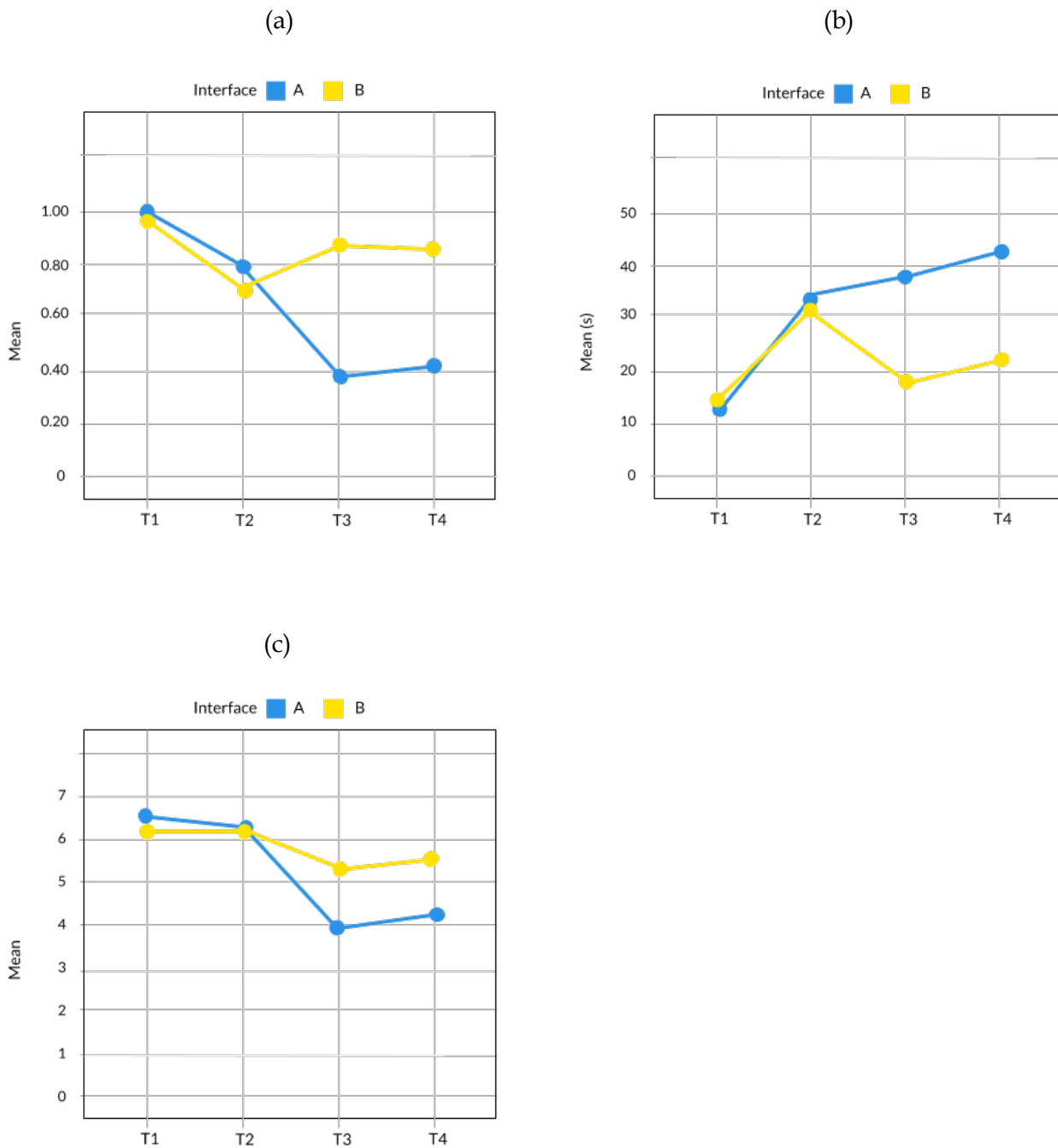


Graphical analysis is usually subjective, and therefore it is also necessary to use more robust statistical tools, such as hypothesis testing. Hence the two-way repeated measures analysis of the variance test was applied to compare the median of each usability metric for each interface by task.

Table 2 shows the main effects of task type and interface type were qualified by a significant interaction between task type and interface type on success rate, completion time and perceived

satisfaction. Due to its interaction effect, a pairwise comparison was conducted to see the simple effects between both factors on the three usability metrics.

*Table 2: p-values when applying the two-way repeated measures ANOVA. Null hypothesis: the median of success rate, completion time and perceived satisfaction of the two interfaces (A: including a tree map versus B: including a stacked ba chart) are equal*

|  | Usability metric | dF | F | p-value |
|---|---|---|---|---|
| Task | Success rate | 3 | 41.395 | .000000* |
|  | Completion time | 3 | 7.421 | .000274* |
|  | Perceived satisfaction | 3 | 4.026 | .011385* |
| Interface | Success rate | 1 | 14.642 | .000321* |
|  | Completion time | 1 | 13.773 | .000464* |
|  | Perceived satisfaction | 1 | 12.001 | .001006* |
| Task * Interface | Success rate | 3 | 6.922 | .000462* |
|  | Completion time | 3 | 6.340 | .000859* |
|  | Perceived satisfaction | 3 | 5.107 | .003328* |
| Error | Success rate | 58 |  |  |
|  | Completion time | 58 |  |  |
|  | Perceived satisfaction | 58 |  |  |

Table 3 presents the p-values from pairwise comparison of the two-way repeated measures ANOVA with Bonferroni's correction. Corroborating and complementing the graphical analysis, comparison tasks (T3 and T4) were significantly different between interfaces in the success rate, in the task completion time and in the perceived satisfaction of the task.

*Table 3: p-values when applying the pairwise comparison to the two-way repeated measures ANOVA test with Bonferroni's correction. *: There are significant differences considering a significance level of 5% and Bonferroni adjustment*

| Task | Usability metric | p-value | Mean difference | |
|---|---|---|---|---|
|  |  |  | (A-B) | (B-A) |
| T1 – Topic categorization | Success rate | 0.3311 | 0.03 | -0.03 |
|  | Completion time | 0.2256 | -1.22 | 1.22 |

| | | | | |
|---|---|---|---|---|
| | Perceived satisfaction | 0.0824 | 0.36 | -0.36 |
| T2 – Opinion categorization | Success rate | 0.0742 | 0.08 | -0.08 |
| | Completion time | 0.0711 | 1.30 | -1.30 |
| | Perceived satisfaction | 0.1345 | 0.16 | -0.16 |
| T3 – Overall comparison | Success rate | 0.0004* | -0.50 | 0.50 |
| | Completion time | 0.0021* | 19.44 | -19.44 |
| | Perceived satisfaction | 0.0244* | -1.33 | 1.33 |
| T4 – Comparison by region | Success rate | 0.0002* | -0.41 | 0.41 |
| | Completion time | 0.0010* | 20.88 | -20.88 |
| | Perceived satisfaction | 0.0285* | -1.41 | 1.41 |

## 5. Discussion

Considering the need for transparency, accountability and public understanding, visualizing public participation data without usability evidence is not enough. On the one hand, recent studies have shown the effectiveness of modeling and visualizing public opinion, but mostly on informal settings and social media (Steinskog, Therkelsen & and Gambäck, 2017; Yu, Xu, Wang & Ni, 2019; Kant, Weisser & Säfken, 2020). On the other hand, the assessment of visualization strategies have often lacked a more comprehensive operationalization of usability and focused mostly on effectiveness.

In this article we examined the usability of two different visual interfaces using data from a participatory constitution-making consultation. We sought to compare the usability of the tree map, which is the most frequently used technique, and the stacked barchart, which is conceptually more apt for quantitative analysis. A two-way repeated measures ANOVA was conducted to determine whether there were significant differences between the two interfaces in terms of success rate, completion time and perceived satisfaction and when being used for two different cognitive tasks (categorization and comparison). This approach was utilized to determine whether the usability of the two interfaces were similar or different when used for solving different tasks.

The results from the ANOVA showed that there is a significant interaction effect between task type and interface in the three dimensions of usability (see Table 2). This suggests that the interfases produce different results but only in certain tasks. In order to find out which interface produces better results in which task, a pairwise comparison was utilized (see Table 3). This analysis showed that the stacked barchart yields better results in all dimensions of usability, but only when solving

tasks 3 and 4, which are comparison tasks. Tasks 1 and 2 (categorization tasks) showed no significant differences.

Our findings support the hypothesis that the stacked barchart is a better alternative to the tree map visualization for cognitive tasks that involve categorization and comparison. This may be explained by the fact that it is more difficult for the human eye to make precise comparisons when the data is encoded through the area of a figure (Few, 2009), as treemap does. On the other hand, the stacked bar chart's advantage may be supported by the hierarchy of visualisation accuracy (Mackinlay, 1986), where the use of position and length to communicate information is the most accurate way to display quantitative data. In this sense, these findings support the idea that the usability of visualization strategies is task-specific and thus, the assessment of public participation modeling and visualization should be directed at specific cognitive operations. More debate and research is needed to determine which cognitive tasks can be used as standards of success for public participation data.

In sum, the contribution of this research is twofold. Firstly, we provide evidence for the usability of two visualization techniques using data from a massive institutional participation exercise. Our results showed that the stacked barchart visualization yields better results for comparison tasks and that categorization tasks show no significant difference. Based on these results, the stacked barchart is more appropriate for visualization that involves both categorizing and comparing quantitative data. Secondly, our research indicates that the performance and usability of visualization strategies are task-specific and thus, that no generic usability assessment should be performed. In turn, this conclusion adds to the complexity of giving back information to the public as visualization strategies have to account for the expected cognitive operations of their final users.

This study focused on the quantitative approach to evaluate the usability of the interfaces. However, the experimental data did not capture the user's subjective experience in performing the tasks. It would be useful to be able to triangulate the experimental results with qualitative information to better understand how the handling of information is represented by the public and how the topic of the information is related to the cognitive task. On the other hand, the participants had a certain level of computer knowledge, since they were all recently graduated professionals. As engineering graduates, they most likely had experience programming visualizations of datasets and were familiar with elements of the task at hand. Previous knowledge and age can influence the results of the experiment. For instance, previous research has highlighted the need to consider educational gaps and digital divides in online participation (Jankowski et al., 2019) Therefore, the study's findings may not be easily generalized to non-computer users and older users. Future research should address how participants characteristics and how the topic of the information impacts the usability of visual interfaces. Finally, these experimental results were not a part of any systematic and mandatory assessment of the participation initiative described. Future studies should address how the assessment of visualizations of public engagement exercises operates as part of larger accountability policies.

### Declaration of interest

The authors declare no conflict of interest.

## Aprendix A

*Table A1: Database extract*

| Argument Id | Region | County | Argument |
|---|---|---|---|
| 45349 | 13 | 307 | There must be equality between men and women. |
| 45235 | 10 | 248 | Men and women should have the same rights and duties for an equitable society. |
| 44861 | 9 | 170 | Gender, rights and opportunities equality must be guaranteed. |

## Aprendix B

*Table B1: Tag and words per topic from topic modeling*

| Topic Id | Topic tag | Words |
|---|---|---|
| 1 | Participation | participation, democracy, decision, consultation, binding, plebiscite |
| 2 | Environment | nature, environment, take, care, resources, conservation |
| 3 | New Constitution | new, constitution, people, defender, constituent, assembly |
| 4 | Congress | congress, parliament, unicameral, representative, democratic, elected |

| 5 | Education and health | education, health, right, access, free, dignified |
| 6 | Decentralization | government, regional, autonomy, local, major, decentralization |
| 7 | Equality before the law | law, equality, justice, norm, all, enforcement |
| 8 | Common good | coexistence, life, good, common, fundamental, respec |
| 9 | Equal rights | all, right, equality, condition, people, discrimination |
| 10 | State Powers | power, judicial, executive, legislative, functions, independent |
| 11 | Cultural Heritage | heritage, cultural, people, protect, our, identity |
| 12 | Human rights | rights, human, protection, promotion, life, essential |
| 13 | Local Economy | economy, free, market, promote, own, resources |
| 14 | Transparency | transparency, positions, public, control, attributions, inspection |
| 15 | Equality and equity | equality, rights, guarantee, equity, gaps, solve |

## Aprendix C

Task-Based Usability Test

*Table C1: Questions by task and version*

| Question Id | Task | Versión 1 | Versión 2 |
|---|---|---|---|
| 1 | Topic categorization | What topic is in the same category as "Common Good"? | What topic is in the same category as "Human Rights"? |
| 2 | Opinion categorization | What topic does the following opinion belong to? "Proactive civic relations are considered essential for the protection of human rights in our society permanently, in order to achieve non-discrimination due to divergent, political and gender opinions" | What topic does the following opinion belong to? "Creation of regional laws and regulations should enable the administration of own resources in a decentralized manner" |
| 3 | Overall comparison | What is the most frequent topic between "Powers of the State" and "Environment"? | What is the most frequent topic between "New Constitution" and Cultural Heritage? |
| 4 | Comparison by region | In Antofagasta region, what is the most frequent topic between "Equality before the Law" and "Participation"? | In Los Lagos region, what is the most frequent topic between "Congress" and "Transparency"? |
| 5 | Topic categorization | What topic is in the same category as "Human Rights"? | What topic is in the same category as "Common Good"? |
| 6 | Opinion categorization | What topic does the following opinion belong to? "Creation of regional laws and regulations should enable the administration of own resources in a decentralized manner" | What topic does the following opinion belong to? "Proactive civic relations are considered essential for the protection of human rights in our society permanently, in order to achieve non-discrimination due |

| | | | to divergent, political and gender opinions" |
|---|---|---|---|
| 7 | Overall comparison | What is the most frequent topic between "New Constitution" and Cultural Heritage? | What is the most frequent topic between "Powers of the State" and "Environment"? |
| 8 | Comparison by region | In Los Lagos region, what is the most frequent topic between "Congress" and "Transparency"? | In the Antofagasta region, what is the most frequent topic between "Equality before the Law" and "Participation"? |

*Figure C1: Example of a topic categorization task*

*Figure C2: Example of an overall categorization task*
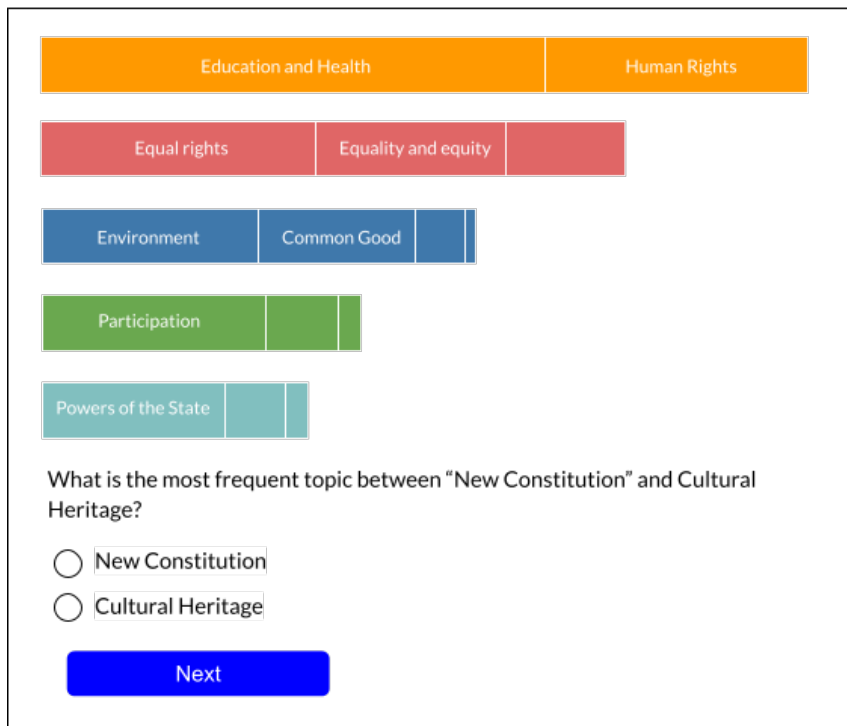


## Apendix D

*Table D1: Means (M) and standard deviation (SD) of success rate, completion time and perceived satisfaction by task type and interface type; A:treemap, B: stacked bac hart (N = 34)*

| Task | Interface | Success rate | | Completion time | | Perceived satisfaction | |
|---|---|---|---|---|---|---|---|
| | | M | SD | M | SD | M | SD |
| Topic categorization | A | 1.00 | 0.00 | 13.43 | 6.10 | 6.58 | 0.53 |
| | B | 0.97 | 0.17 | 14.65 | 4.76 | 6.22 | 0.56 |
| Opinion categorization | A | 0.79 | 0.41 | 33.03 | 10.72 | 6.29 | 0.98 |
| | B | 0.71 | 0.46 | 31.73 | 11.36 | 6.13 | 0.95 |
| Overall comparison | A | 0.38 | 0.49 | 38.29 | 8.90 | 3.92 | 1.16 |
| | B | 0.88 | 0.33 | 18.85 | 3.92 | 5.25 | 0.58 |
| Comparison by region | A | 0.44 | 0.50 | 43.13 | 7.24 | 4.12 | 0.67 |
| | B | 0.85 | 0.36 | 22.25 | 5.68 | 5.53 | 0.88 |

# References

Arnstein, S. R. (1969). A ladder of citizen participation. *Journal of the American Institute of planners*, 35(4), 216-224.

Barnum, C. M. (2020). *Usability testing essentials: ready, set... test!*. Morgan Kaufmann.

Baum, H. S. (2015). Citizen participation. In: N.J. Smelser and P.B. Baltes, eds. *International encyclopedia of the social and behavioral sciences*. New York: Elsevier, 1840–1846

Becker Castellaro, S. (2018). Análisis deliberativo a la participación en procesos constituyentes: El caso chileno (2015-2018).

Bertin, J. (1983). *Semiology of graphics; diagrams networks maps* (No. 04; QA90, B7.).

Bobbio, L. (2019). Designing effective public participation. *Policy and Society*, *38*(1), 41–57. https://doi.org/10.1080/14494035.2018.1511193

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.

Brynjolfsson, E., & McAfee, A. (2017). What's driving the machine learning explosion. *Harvard Business Review,* 18.

Cai, G., Sun, F., & Sha, Y. (2018, January). Interactive Visualization for Topic Model Curation. In *IUI Workshops*.

Calheiros, A. C., Moro, S., y Rita, P. (2017). Sentiment classification of consumer-generated online reviews using topic modeling. *Journal of Hospitality Marketing & Management*, 26(7), 675-693.

Cardullo, P., & Kitchin, R. (2019). Being a 'citizen in the smart city: up and down the scaffold of smart citizen participation in Dublin, Ireland. *GeoJournal*, 84(1), 1-13.

Chaney, A. J. B., y Blei, D. M. (2012). Visualizing topic models. In  *Sixth International AAAI Conference on Weblogs and Social Media*-ICWSM, pp. 419-422, 2012.

Cleveland, W. S., & McGill, R. (1985). Graphical perception and graphical methods for analyzing scientific data. *Science*, 229(4716), 828-833.

Correll, M., & Gleicher, M. (2014). Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE transactions on visualization and computer graphics*, 20(12), 2142-2151.

Corry, M. D., Frick, T. W., & Hansen, L. (1997). User-centered design and usability testing of a web site: An illustrative case study. *Educational technology research and development*, 45(4), 65-76.

Demiralp, Ç., Bernstein, M. S., & Heer, J. (2014). Learning perceptual kernels for visualization design. *IEEE transactions on visualization and computer graphics*, 20(12), 1933-1942.

DePuy, V., & Berger, V. W. (2014). Counterbalancing. *Wiley StatsRef: Statistics Reference Online.*

Dimara, E., Bezerianos, A., & Dragicevic, P. (2017). Conceptual and methodological issues in evaluating multidimensional visualizations for decision support. *IEEE transactions on visualization and computer graphics*, 24(1), 749-759.

Dryzek, J. S. (2002). *Deliberative democracy and beyond: Liberals, critics, contestations*. Oxford: Oxford University Press.

Eisenstadt, T. A., LeVan, A. C., & Maboudi, T. (2015). When talk trumps text: the democratizing effects of deliberation during constitution-making, 1974-2011. *The American Political Science Review*, 109(3), 592.

Elmqvist, N., & Yi, J. S. (2015). Patterns for visualization evaluation. *Information Visualization*, 14(3), 250-269.

Few, S. (2009). *Now you see it: simple visualization techniques for quantitative analysis* (No. Sirsi) i9780970601988).

Fierro, C., Fuentes, C., Pérez, J., & Quezada, M. (2017). 200k+ crowdsourced political arguments for a new chilean constitution. In *Proceedings of the 4th Workshop on Argument Mining* (pp. 1-10).

Floridia, A. (2018). The Origins of the Deliberative Turn. In A. Bächtiger, J. S. Dryzek, J. Mansbridge, & M. Warren (Eds.), *The Oxford Handbook of Deliberative Democracy* (pp. 34–54). Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198747369.013.25

Ganesan, A., Brantley, K., Pan, S., & Chen, J. (2015). Ldaexplore: Visualizing topic models generated using latent dirichlet allocation. *arXiv preprint arXiv*:1507.06593.

Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive psychology*, 1(3), 225-241.

Gluck, J., & Ballou, B. (2014). *New Technologies in Constitution Making*. Washington, D.C.

Gluck, J., & Brandt, M. (2015). *Participatory and inclusive constitution-making*. Washington, D.C.

Hagen, L., Uzuner, Ö., Kotfila, C., Harrison, T. M., y Lamanna, D. (2015). Understanding citizens' direct policy suggestions to the federal government: a natural language processing and topic modeling approach. In *2015 48th Hawaii International Conference on System Sciences* (pp. 2134-2143). IEEE.

Howorko, L., Boedianto, J. M., & Daniel, B. (2018). The efficacy of stacked bar charts in supporting single-attribute and overall-attribute comparisons. *Visual Informatics*, 2(3), 155-165.

Hsiao, Y. T., Lin, S. Y., Tang, A., Narayanan, D., y Sarahe, C. (2018). vTaiwan: An empirical study of open consultation process in Taiwan.

Humphreys, A., & Wang, R. J. H. (2018). Automated text analysis for consumer research. *Journal of Consumer Research*, 44(6), 1274-1306.

Irvin, R. A., & Stansbury, J. (2004). Citizen participation in decision making: is it worth the effort?. *Public administration review*, 64(1), 55-65.

Isaac, O., Abdullah, Z., Ramayah, T., & Mutahar, A. M. (2017). Internet usage, user satisfaction, task-technology fit, and performance impact among public sector employees in Yemen. *The International Journal of Information and Learning Technology*.

ISO. (1998). ISO 9241-11:1998 Ergonomic requirements for office work with visual display terminals (vdts) – part 11: guidance on usability.

Jankowski, P., Czepkiewicz, M., Młodkowski, M., Zwoliński, Z., & Wójcicki, M. (2019). Evaluating the scalability of public participation in urban land use planning: A comparison of Geoweb methods with face-

to-face meetings. *Environment and Planning B: Urban Analytics and City Science*, *46*(3), 511–533. https://doi.org/10.1177/2399808317719709

Jankowski, P., & Nyerges, T. (2001). GIS-supported collaborative decision making: Results of an experiment. Annals of the Association of American Geographers, 91(1), 48-70.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet Allocation (LDA) and Topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78(11), 15169-15211.

Kant, G., Weisser, C., & Säfken, B. (2020). TTLocVis: A Twitter Topic Location Visualization Package. *Journal of Open Source Software*, 5(54), 2507.

Karpovich, S., Smirnov, A., Teslya, N., & Grigorev, A. (2017, April). Topic model visualization with ipython. In *2017 20th Conference of Open Innovations Association (FRUCT)* (pp. 131-137). IEEE.

Keller, P. R., Keller, M. M., Markel, S., Mallinckrodt, A. J., & McKay, S. (1994). Visual cues: practical data visualization. *Computers in Physics*, 8(3), 297-298.

Kim, Y., & Heer, J. (2018, June). Assessing effects of task and data distribution on the effectiveness of visual encodings. In *Computer Graphics Forum* (Vol. 37, No. 3, pp. 157-167).

Kosara, R. (2019). The Impact of Distribution and Chart Type on Part-to-Whole Comparisons. In *EuroVis (Short Papers)* (pp. 7-11).

Kong, N., Heer, J., & Agrawala, M. (2010). Perceptual guidelines for creating rectangular treemaps. *IEEE transactions on visualization and computer graphics,* 16(6), 990-998.

Lewis, J. R. (1995) IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. *International Journal of Human-Computer Interaction*, 7:1, 57-78.

Lewis, J. R. (Ed.). (2019). *Usability Evaluation: A Special Issue of the International Journal of Human-Computer Interaction*. CRC Press.

Li, G., Zhang, Y., Dong, Y., Liang, J., Zhang, J., Wang, J., & Yuan, X. (2019). BarcodeTree: Scalable comparison of multiple hierarchies. *IEEE transactions on visualization and computer graphics*, 26(1), 1022-1032.

Liu, Q., Chen, Q., Shen, J., Wu, H., Sun, Y., & Ming, W. K. (2019). Data analysis and visualization of newspaper articles on thirdhand smoke: a topic modeling approach. *JMIR medical informatics*, 7(1), e12414.

Long, L. K., Hui, L. C., Fook, G. Y., & Zainon, W. M. N. W. (2017). A Study on the Effectiveness of Tree-Maps as Tree Visualization Techniques. *Procedia Computer Science*, 124, 108-115.

Maboudi, T., & Nadi, G. P. (2016). Crowdsourcing the Egyptian constitution: social media, elites, and the populace. *Political Research Quarterly*, 69(4), 716-731.

Mackinlay, J. (1986). Automating the design of graphical presentations of relational information. *Acm Transactions On Graphics (Tog)*, 5(2), 110-141.

Manresa-Yee, C., Amengual, E., & Ponsa Asensio, P. (2014). La usabilidad de las interfaces basadas en visión. *FAZ*, (7), 12-31.

Mayr, S., Erdfelder, E., Buchner, A., & Faul, F. (2007). A short tutorial of GPower. *Tutorials in quantitative methods for psychology*, 3(2), 51-59.

Munzner, T. (2009). A nested model for visualization design and validation. *IEEE transactions on visualization and computer graphics,* 15(6), 921-928.

Mukherjee, S. (2019). How IT allows E-Participation in Policy-Making Process. *arXiv preprint arXiv*:1903.00831.

Negretto, G. L. (2017). Constitution Making in Democratic Constitutional Orders: The Challenge of Citizen Participation. *Let the people rule*.

Noveck, B. S. (2018). Crowdlaw: Collective intelligence and lawmaking. *Analyse & Kritik*, 40(2), 359-380.

OECD (2017). Public Governance Reviews. Chile scan report on the citizen participation in the constitutional process. Technical report.

Ogao, P. J., & Kraak, M. J. (2002). Defining visualization operations for temporal cartographic animation design. *International journal of applied earth observation and geoinformation*, 4(1), 23-31.

Pacheco i Canals, J. (2019). A normative model for deliberative constitution-building. *Revista Catalana de Dret Públic*, *58*, 166–181. https://doi.org/http://dx.doi.org/10.2436/rcdp.i58.2019.3225

Parra, D., Trattner, C., Gómez, D., Hurtado, M., Wen, X., & Lin, Y. R. (2016). Twitter in academic events: a study of temporal usage, communication, sentimental and topical patterns in 16 computer science conferences. *Computer Communications*, 73, 301-314.

Polasanapalli, S. L., & Buggareddy, P. (2020). Usability Evaluation to design a user interface by implementing HCI design principles.

Qian, L., Wachowicz, M., Peuquet, D., & MacEachren, A. (1997, October). Delineating operations for visualization and analysis of space-time data in GIS. In *Gis/lis* (Vol. 97, No. October, pp. 28-30).

Qu, Z., & Hullman, J. (2016, October). Evaluating visualization sets: Trade-offs between local effectiveness and global consistency. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization* (pp. 44-52).

Radford, A., Narasimhan, K., Salimans, T., y Sutskever, I. (2018). Improving language understanding by generative pre-training. Retrieved from www.s3-us-west-2.amazonaws.com/openai-assets/research-covers/languageunsupervised/language understanding paper.pdf.

Raveau, M. P., Couyoumdjian, J. P., Fuentes-Bravo, C., Rodriguez-Sickert, C., & Candia, C. (2020). Citizens at the forefront of the constitutional debate: Participation determinants and emergent content in Chile. *arXiv preprint arXiv*:2006.00140.

Roque de Oliveira, A., & Partidário, M. (2020). You see what I mean? – A review of visual tools for inclusive public participation in EIA decision-making processes. *Environmental Impact Assessment Review*, *83*, 106413. https://doi.org/10.1016/j.eiar.2020.106413

Saket, B., Endert, A., & Demiralp, Ç. (2018). Task-based effectiveness of basic visualizations. *IEEE transactions on visualization and computer graphics*, 25(7), 2505-2512.

Sørum, H., Andersen, K. N., & Vatrapu, R. (2012). Public websites and human–computer interaction: an empirical study of measurement of website quality and user satisfaction. *Behaviour & Information Technology*, 31(7), 697-706.

Steinskog, A., Therkelsen, J., & Gambäck, B. (2017, May). Twitter topic modeling by tweet aggregation. In *Proceedings of the 21st nordic conference on computational linguistics* (pp. 77-86).

Sukhija, N., Tatineni, M., Brown, N., Van Moer, M., Rodriguez, P., & Callicott, S. (2016, July). Topic modeling and visualization for big data in social sciences. In *2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress* (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld) (pp. 1198-1205). IEEE.

Sun, S., Luo, C., & Chen, J. (2017). A review of natural language processing techniques for opinion mining systems. Information fusion, 36, 10-25.

Szafir, D. A., Haroz, S., Gleicher, M., & Franconeri, S. (2016). Four types of ensemble coding in data visualizations. *Journal of vision*, 16(5), 11-11.

Umar, M. M., Bakhat, M. U., & Hassan, M. (2020). Mapping HCI Principles to Evaluate the Usability of Learning Applications for CCI Users. *International Journal of Computer Science and Telecommunications*, 11(2).

Vernier, E. F., Telea, A. C., & Comba, J. (2018, September). Quantitative comparison of dynamic treemaps for software evolution visualization. In *2018 IEEE Working Conference on Software Visualization* (VISSOFT) (pp. 96-106). IEEE.

Ware, C. (2019). *Information visualization: perception for design*. Morgan Kaufmann.

Wehrend, S., & Lewis, C. (1990, October). A problem-oriented classification of visualization techniques. In *Proceedings of the First IEEE Conference on Visualization: Visualization 90* (pp. 139-143). IEEE.

Xu, Z., Liu, Y., Xuan, J., Chen, H., & Mei, L. (2017). Crowdsourcing based social media data analysis of urban emergency events. *Multimedia Tools and Applications*, 76(9), 11567-11584.

Yang, Y., Yao, Q., y Qu, H. (2017). VISTopic: A visual analytics system for making sense of large document collections using hierarchical topic modeling. *Visual Informatics*, 1(1), 40-47.

Yu, D., Xu, D., Wang, D., & Ni, Z. (2019). Hierarchical topic modeling of Twitter data for online analytical processing. *IEEE Access*, 7, 12373-12385.

Zia, A., Guo, L., Zhou, L., Essa, I., & Jarc, A. (2019). Novel evaluation of surgical activity recognition models using task-based efficiency metrics. *International journal of computer assisted radiology and surgery*, 14(12), 2155-2163.

Zhou, M. X., & Feiner, S. K. (1998, January). Visual task characterization for automated visual discourse synthesis. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 392-399).

## About the Author

*Ivania Yovanovic*

Ivania Yovanovic is an industrial engineer interested in data visualization, open government and public in-novation. As a public employee for the Chilean Government, she works on visualization projects of public data.

*Julian Goñi*

Julian Goñi is a psychologist specialized in interdisciplinary research. As a lecturer and researcher at the Engineering Design Initiative, Universidad Católica de Chile (DILAB UC), he researches the intersection be-tween engineering design, social research and responsible innovation.

*Constanza Miranda*

Constanza Miranda is a multidisciplinary academic interested in the intersection between the creativity of design, the ethics of cultural anthropology and the tech aspects of engineering. With a PhD in Design from NCSU she is now faculty at Johns Hopkins BME department and does research in design education.