

# New Findings on the User's Preferences about Data Visualization of Online Reviews

Elizabeth Simão Carvalho, Marcirio Silveira Chaves

**Abstract**—The information visualization is still a knowledge field that lacks from a solid theory to support it and there is a myriad of existing methodologies and taxonomies that can be combined and adopted as guidelines. In this context, it is necessary to pre-evaluate as much as possible all the assumptions that are considered for its design and development. We present an exploratory study ( $n = 123$ ) to detect the graphical preferences of travelers using accommodation portals of Web 2.0 (e.g. tripadvisor.com). We took into account some of the most relevant ground rules applied in the field to map visually data and design end-user interaction. Moreover, the evaluation process was completely data visualization oriented. We found out that people tend to refuse more advanced types of visualization and that a hybrid combination between radial graphs and stacked bars should be explored. In sum, this paper introduces new findings about the visual model and the cognitive response of users of accommodation booking websites.

**Keywords**—Information visualization, Data visualization, Visualization evaluation, Online reviews, Booking portal, Hotel booking.

## I. INTRODUCTION

INFORMATION visualization techniques offer a more varied palette of visual characteristics than the typical interface: Characteristics such as color, brightness, contrast, surface area, line weight, and visual pattern, among many others, can be used to represent differences between pieces of data. Indeed, one of the great challenges of information visualization is to know when to say no to this embarrassment of riches – the hardest choices are often among the few visual aspects that will best represent the underlying data set in a way that's easy to grasp.

Research about data visualization applied to opinions freely available on the web has received more attention in the last years (Gamon et al [1]; Draper and Riesenfeld [2]; Wu et al [3]). This can be due to the fact that people are increasingly using websites to find out information about potentially anything. For instance, it is becoming a global habit to visit accommodation portals before booking hotels. Almost all respondents of a survey performed by Gretzel et al [4] use the Internet as an information source for planning pleasure trips and most (82.5%) uses it every time they plan a pleasure trip.

Most of accommodation websites have the particularity of holding and displaying reviews generated by the hosts. This

web content, although being helpful to support the traveler decision when booking a room, is usually displayed as a text list, taking not into account any visual, interaction or cognitive aspect of the website end-user.

On the other hand, this data is generally poorly treated and analyzed before being made available to the website visitor. This means that much more information may be extracted and inferred from the data if it is properly processed. For example, the Dayton Hudson Corp. success story in the retail market shows how the use of visual data mining helped them grow their business and improve customer satisfaction, while the Sprint success story in the telecommunication market demonstrates how it enhanced the customer relationship management. These and many other good examples of data visualization success are cited by Soukup and Davidson [5].

Although the use of advanced graphic techniques in online reviews is not recent, preliminary advanced graphical representation of online reviews in the accommodation sector was firstly proposed in Carvalho and Chaves [6]. In this paper, we deepen this study [6] through a survey on information visualization of online reviews about the accommodation sector. The main goals of this survey are:

- To pursue towards understanding the work, analysis, and information processing practices associated with the consulting of accommodation booking information;

- To elicit formal requirements for design;

- To know what target users think of the different visualization techniques and what features are seen as useful;

The answers are further analyzed in order to identify and correlate possible patterns in user's profiles and visual and interaction attributes. The major objective to be achieved is the identification of a visual model of the potential end-user that should be considered to implement the data visualization component.

To the best of our knowledge, this is the first work to present findings about the visual model and the cognitive response of users of accommodation booking websites.

This paper is organized into five sections. The background section describes previous works related to accommodation sector data visualization and evaluation methodologies that were taken into account to model the visualization process and design and evaluate the survey. The work description section is divided into two other sections: Sample data characterization and data analysis. It gives a detailed overview about how this survey was developed, the data considered and how it was treated, and the measured numerical results. The discussion section launches conclusions deduced from the results achieved. The findings section highlights the features

Elizabeth S. Carvalho is with University Aberta, Palácio Ceia, rua da Escola Politécnica, nº 141-147, 1269-001 Lisbon, Portugal (phone: + 351 213 916 300; e-mail: elizabeth.carvalho@uab.pt).

Marcirio S. Chaves is with Universidade Nove de Julho – UNINOVE, Av. Francisco Matarazzo, 612 Building C, first floor. Água Branca 05001-000 São Paulo, Brazil (e-mail: mschaves@uninove.br).

that should be present in a visual model considered useful by users and the final remarks points out the main conclusions of this work, limitations and future works.

## II. BACKGROUND

This work was developed as a combination of different scientific background. Data visualization and its usage in the scope of sentiment and opinion analysis of the blogosphere was reviewed in order to find out what kind of visual

representations are being mostly considered. On the other hand, it was also important to identify methodologies that would be used to guide the design process of data visualization and its evaluation. Finally, because this work is related to online reviews of the accommodation sector, it was also considered relevant to get a better insight of the travelers' profile and their preferences.

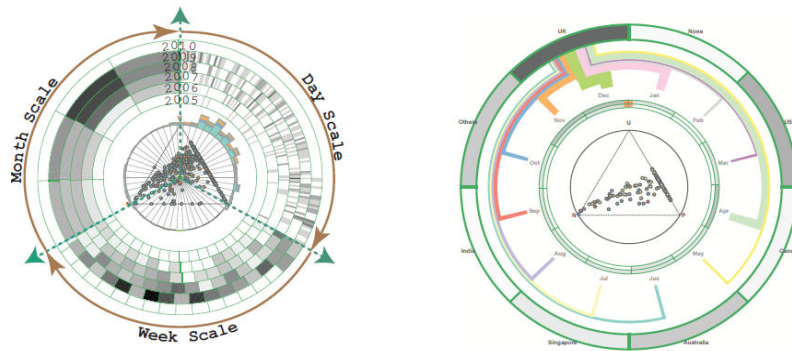


Fig. 1 Temporal rings at different scales and temporal and geographic rings [3]

### A. Online Reviews Data Visualization

Gamon et al [7] developed a system named Pulse, which contains a user interface based on the Tree Map technique to represent features from sentences in online reviews. The Pulse prototype system mines topics and sentiment orientation jointly from free text customer feedback.

Wu et al [3] designed and developed OpinionSeer to address the need to effectively communicate opinion-mining results and facilitate the analytical reasoning process. The system was built on a new visualization-centric opinion mining technique that considers uncertainty for faithfully modeling and analyzing customer opinions. A new visual representation was developed to convey customer opinions by augmenting well-established scatterplots and radial visualization (Fig. 1 illustrates some outputs). In Fig. 1, the opinion wheel seamlessly integrates a scatterplot (opinion triangle) with a radial visualization (opinion ring). The opinion triangle is primarily used for visualizing the extracted opinions while the opinion rings surrounding the triangle facilitate the visual exploration of correlations between the customer opinions and other data dimensions. The opinions in the triangle are projected onto the opinion rings to create circular histograms of different data dimensions.

Draper and Riesenfeld [2] presented a simple interactive visualization that allows users to construct queries on large tabular datasets from surveys and opinion polls, and view the results in real time. The results of two separate user studies suggested that their interface lowers the learning curve for naive users, while still providing enough analytical power to discover interesting correlations in the data.

Gamon et al [1] tracked thousands of blogs and the news articles they cite, collapsing news articles that have highly

overlapping content. They then tag each article with the number of blogs citing it, the political orientation of those blogs, and the level of emotional charge expressed in the blog posts that link to the news article. They summarized and presented the results to the user via a novel form of visualization. The BLEWS system is a first foray into using a combination of text processing and link analysis to provide contextual information for news articles as required.

### B. Visualization Design and Evaluation Methodology

Lam et al [8] proposed seven scenarios to evaluate visualization that were derived through an extensive literature review of over 800 publications. In the scope of our work, we considered that two of these scenarios should be used to guide the process of visualization design and evaluation: the EWP - Evaluating Environments and Work Practices and the EU - Evaluating User Experience.

The EWP helps to elicit formal requirements for design. The goal of information visualization evaluations in this category is to pursue towards understanding the work, analysis, for information processing practices by a given group of people with or without software in use. The output of studies in this category are often design implications based on a more holistic understanding of current workflows and work practices, the conditions of the working environment itself, and potentially current tools in use.

The EU seeks to understand how people react to visualization either in a short or a long time span. Visualization here may interchangeably be intended as an initial design sketch, a working prototype, as well as a finished product. The goal is to understand to what extent the visualization supports the intended tasks as seen from the participants' eyes and to probe for requirements and needs.

In terms of design, visualization design requires taxonomy of techniques to guide automated algorithms, programmers, or users. It is an area still under debate and evolution. Most visualization taxonomies are based on the type of data involved [9]. Data has many characteristics, including: Number of independent variables, number of dependent variables, type of each variable (e.g. scalar, vector, discrete, or nominal). Tweedie [10] describes several forms of data that may be represented visually: data values (described above), data structure (e.g. file hierarchies, or rectilinear vs. curvilinear grids), and metadata.

“Information visualization” can be similarly organized by data type [11]. Common categories are multi-dimensional databases (often containing more than three dimensions), text, graphs, and trees [12]. In addition to data type, some taxonomies [13] have organized visualization systems by display style (e.g. table or information landscape) or include generic tasks performed by users of the system (e.g. gaining an overview, drilling down on details or filtering).

### C. Traveler's Profile and Online Booking Behavior

A better knowledge of the traveler's profile is important to have a clearer picture of what his visual model potentially may be. Other relevant aspect is to know his accommodation booking behavior and thus, what kind of information is more important when he interacts with a travelling website. This information identifies possible measures and dimensions to be made available to the end-user in the data visualization interface.

Gretzel et al [4] did a study with five specific goals. The survey was administered during a 4 week period between

January 5 and January 31, 2007. The respondents were from a TripAdvisor.com traveler panel. The questions were related to general travel experience/planning style, perceptions, use and impact of online travel reviews, motivations and barriers to posting online reviews, and included demographic information. Some of the main conclusions they achieved were:

Almost all respondents use the Internet as an information source for planning pleasure trips and most (82.5%) uses it every time they plan a pleasure trip;

Most online travel review readers indicate that detailed descriptions (71%), the type of Web site where the review is posted (64.7%), and the date the review was posted (59.3%) are critical to evaluating a review;

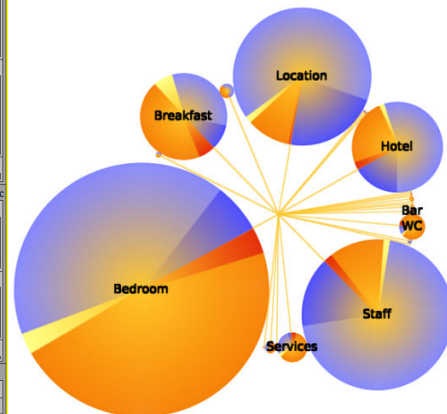
A reviewer's credibility is most frequently judged based on the reviewer's travel experience (75.3%), similarity in terms of activities engaged in during a trip (65.9%), trip purpose (60%) and writes in a polite and friendly manner (58.5%);

Similarly, a review's usefulness is most frequently judged based on the reviewer's travel experience (78.5%), similarity in terms of activities engaged in during a trip (68.2%), trip purpose (64.4%) and similarity in terms of age, gender or marital status (53.1%).

In short, this means that measures such as the reviewer's experience, the similarity of the activity engaged during the trip and its purpose or the clarity of the comment left, are highly taken into account to consider or not the reviewer's opinion. The age, gender, marital status and the date the review was posted are dimensions that should be made available, although not having the same importance.



(a)



(b)

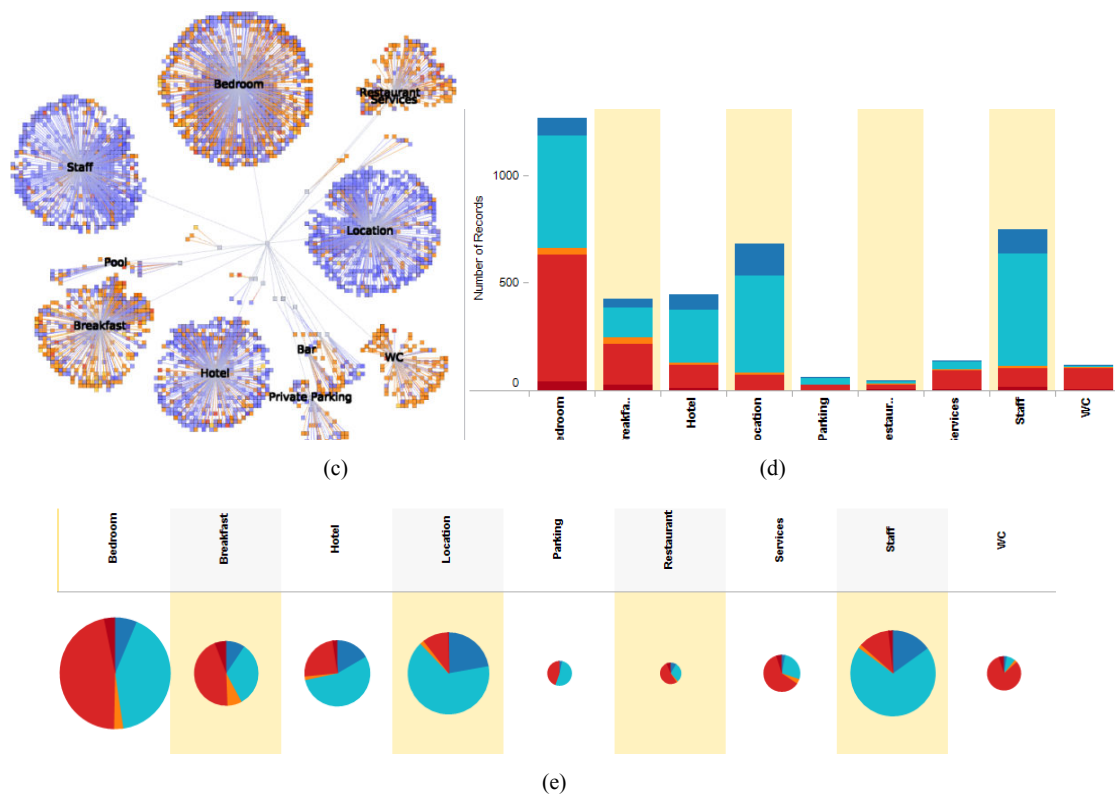


Fig. 2 The five graphics used in the questionnaire – (a) treemap, (b) radial graph, (c) force-directed, (d) stacked bar and (e) pie charts

### III. METHOD AND PROCEDURE

This research presents a work developed having as background user-generated data, specifically to the accommodation sector. Online reviews collected from guests of small and medium hotels (SMH) were used and cross-referenced to find patterns that could give a better insight and support an effective decision making. To classify and analyze the data, a methodology based on concepts of a domain-specific ontology — Hontology, described in detail in Chaves et al. [14] was applied.

A 25-item questionnaire was developed and pre-tested on a sample of 29 people to ensure clarity and feasibility. It was divided in four main sections. Each of them was focused on an aspect that we considered important to identify in the end-user visual model: visual attributes, interaction attributes, dimensions and measures of data, and the end-user profile.

Based on the related work described in the Section Online reviews data visualization, we decided that our data should be presented to the subjects in the survey according to a tree-oriented visualization, besides presenting it in more trivial graph layouts (such as the bar (d) and pie (e) charts illustrated in Fig. 2). To achieve this, we clustered the data and identified the existing relationships. Then, we plotted a squarified treemap (a), a radial (b) and a force-directed (c) graphs, mapping the relationship between the traveler profile, the ontology and polarity strength (e.g. very positive and very

negative). After a brief textual contextualization, respondents face five graphics to evaluate, as depicted in Fig. 2.

The evaluation of these graphs demanded that the subjects chose and indirectly voted what visual variables were more relevant to them. Each of these graphs offers a slightly different cognitive perception to the end-user. The questions made upon them, tried to take advantage of this and identify the most suited visual attributes.

Although the number of questions in the interaction attributes section to be less than the visual one, we tried to concentrate our questions on finding out if data exploration tasks based on direct interaction or visual cues change would be an asset to the end-user. The same shortcut we had in terms of identifying the most desired dimensions and measures of the data to be visually represented and the end-user profile. We could not include so many questions as we would like because the overall survey should meet a good ratio between number of questions and time spent in answering. We assumed that the fulfillment of the whole survey should not exceed 15 minutes; otherwise our subjects would start to feel bored. In our pre-test this premise was upheld.

Finally, participants received an e-mail containing a link to the questionnaire web site. The survey was administered during a six week period between April 2 and May 15, 2012.

#### A. Data Analysis

The Web-based survey effort led to 123 useable responses, which represents a 21 percent response rate. Mean age was 38

(SD = 9 years) with the largest portion (47.8%) falling between the ages of 31 to 40 years. Most of respondents (84%) had at least college degree and 67 percent were man. Eighty-three percent of the respondents usually read the reviews written in booking websites, while 77 percent of the respondents consider that read the reviews is decisive or crucial to book an accommodation. Regarding to the nationality, 57 percent of the respondents lived the most time of their life in Portugal and 38.5 percent in Brazil. For a complete profile of the sample characteristics, see the appendix.

### B. Visual Attributes

Nine questions were focused on identifying what visual attributes should be present in the redesigned version of the data visualization about the accommodation sector. The goals of questions 1 up to 8 are summarized in Table I, while in Fig. 3 a bar graph illustrates the frequency of the most voted chart in each question.

TABLE I  
MOST VOTED GRAPHICS FOR EACH VISUAL ATTRIBUTE

Question number	Question goal (Identify what chart...)	Number of the most voted chart
Q1	is easier to compare values	G4
Q2	is easier to locate information without losing the general context	G2
Q3	is easier to obtain a summary overview of the information	G4
Q4	has the color as the most relevant visual attribute to map information	G2
Q5	has the shape as the most relevant visual attribute to map information	G2
Q6	has the position (spatial location) as the most relevant visual attribute to map information	G2
Q7	is easier to quantify the number of comments about a certain aspect	G4
Q8	offer the better insight of the comments about the different aspects of the accommodation	G4

The most voted charts were 2 and 4. Chart 2 was a radial graph, which is specially used when the hierarchy or existing dependency in data needs to be emphasized. In terms of relevant perceptual codification, this graph explores mostly the position, the size, the proximity, the connectivity, the shape (circular) and the color as visual attributes to map data.

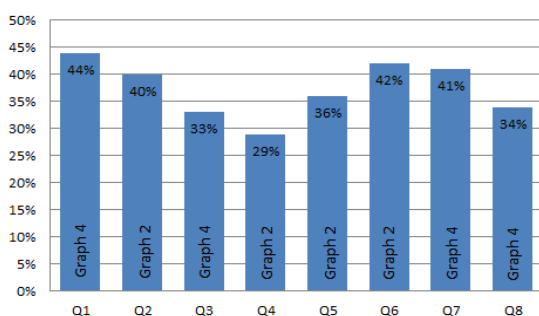


Fig. 3 Resulting frequencies to questions 1-8

Chart 4 was a regular stacked bar chart which is normally used when we need to emphasize quantitative differences between data. It cannot map so much information as a radial graph, although it offers a simpler visualization of data. Moreover, it has been around for a long time, which makes it readable for almost everyone. Its significant perceptual codifications are the position, the size, the color and the shape (rectangular).

According to Table I, 50 percent of the questions pointed out graph 2 as being the best to locate information without losing the general context or as having better used the shape visual attribute, for instance. Meanwhile, 50 percent of the questions pointed out graph 4 as being more suitable in other aspects. Another result achieved, was that mostly women voted in the graph 4, while mostly men voted at graph 2.

The other question that aimed at identifying visual attributes was specially centered at the color issue. Considering that color is one of the most important visual cues, we think that to allow the subject to identify what color palette should be used to map data from online reviews from the accommodation sector, should be included in the survey. In this question, nine colors were made available to choose (including the primary ones) and order, besides the possibility of adding a comment suggesting other ones. The answers to these questions were analyzed according two perspectives: in what order colors are preferred and what combination is most suitable.

Fig. 4 illustrates the results achieved. The blue and green colors were the most voted for the first position (64% and 40%, respectively); while the yellow was the most voted for the second and third position (37% and 29%, respectively). In terms of color combination, we did not achieve a clear result because most of the subjects chose all the colors (15%), except the gray one. Following this result, the most significant result (9.2%) indicated that a combination between green, red, yellow, blue and orange colors should be used to map the information conveniently. Figs. 5 and 6 show other perspectives achieved with the color issue. In Fig. 5, the green and red colors were the mostly voted by subjects with ages between 26 and 45. The blue and yellow ones appear slightly more significant than green only for the group with ages between 31 and 35. Subjects with ages between 51 and 63 tend to choose yellow or blue as preferred colors. In terms of gender, we can see clearly that men were the most responsible for the red and green being chosen according to Fig. 6.



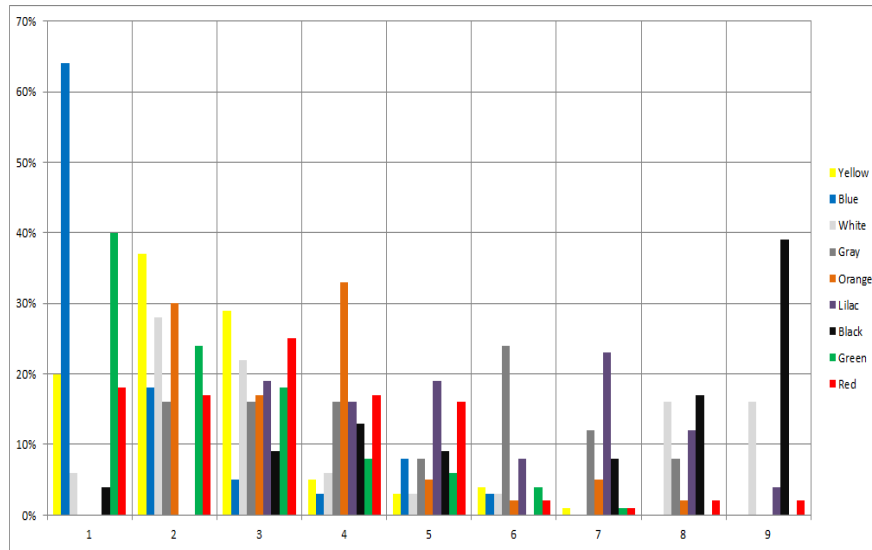


Fig. 4 Color palette results sorted by position (1-9) in which they appear in the user's preference

Figs. 5 and 6 show other perspectives achieved with the color issue. In Fig. 5, the green and red colors were the mostly voted by subjects with ages between 26 and 45. The blue and yellow ones appear slightly more significant than green only for the group with ages between 31 and 35. Subjects with ages

between 51 and 63 tend to choose yellow or blue as preferred colors. In terms of gender, we can see clearly that men were the most responsible for the red and green being chosen according to Fig. 6.

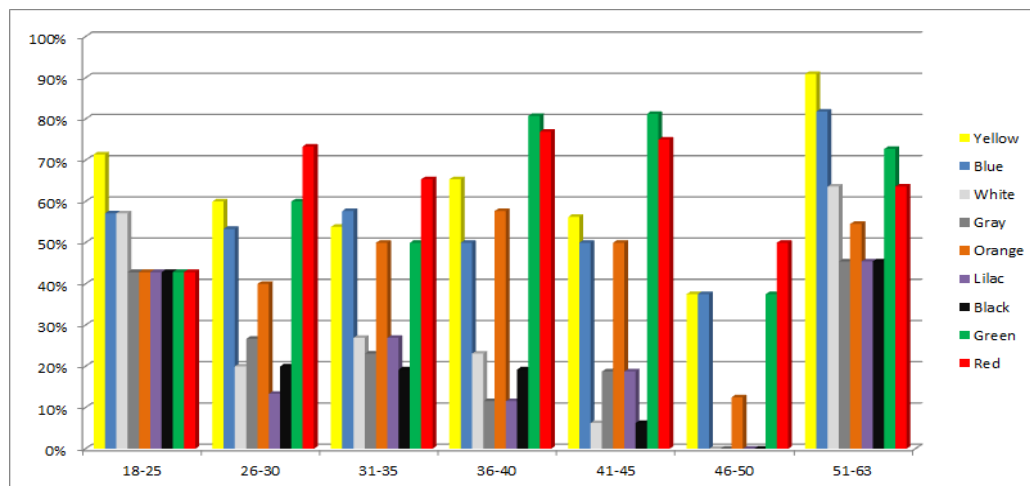


Fig. 5 Color palette results according age clusters

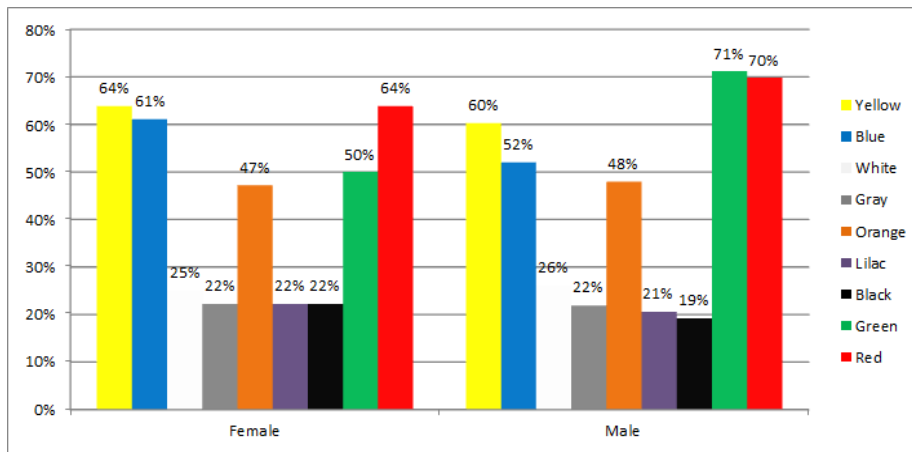


Fig. 6 Color palette results according gender clusters

### C. Dimensions and Measures

The data considered is composed and defined by several variables. Some of them, are for example, the name of the hotel, its number of stars and rating, the facilities that are offered by it, the concept of ontology and its strength or the traveler travelling condition (if alone or in a group, with or without family, etc.) and his origin besides others. These variables define potential dimensions to be used to create the visual representations of the data. Most of these dimensions are likely to be measured somehow or even be classified as measures itself. For instance, the average of the rating or the maximum and minimum strengths achieved by a concept of ontology would be interesting to be made available to the end-user in the visual representations as measures. The number of stars can be considered both as a dimension (ordinal) and a measure itself.

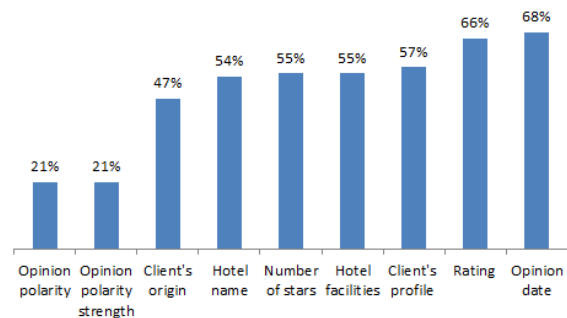


Fig. 7 Distribution of frequencies by preferred graph dimensions

As mentioned in Section III, one of our concerns in the survey was to find out as much as possible, what dimensions and measures should be included in the visual mapping of the data. To evaluate this, we included a multiple choice question where the user should choose what dimensions would be offered to him and also considered mostly relevant to support decision. As done before, it was also included a field to receive the subjects comments and suggestions about other dimensions to be considered. Figs. 7 and 8 illustrate some of the results achieved.

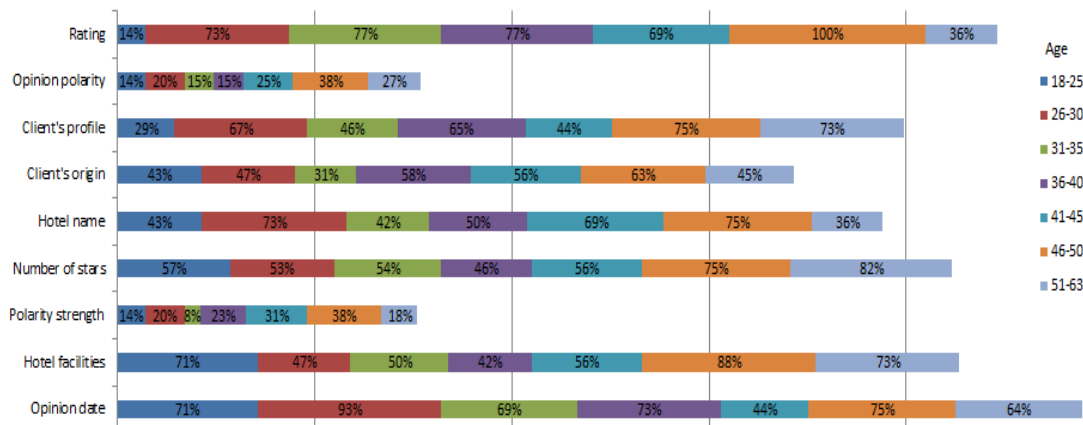


Fig. 8 Dimensions choice frequency versus subjects age scale

Fig. 7 indicates that the opinion date and rating were the most important dimensions to be included in the visualization. In the following, the traveler's profile and the number of stars and the hotel facilities were also considered relevant. Fig. 8 shows how the dimensions were chosen according to the subjects age scale. The rating dimension (or measure) was mostly voted by the subjects with ages between 46 and 50 (100%), followed by those with ages between 31 and 35 and 36 and 40 (each with 77%). The number of stars was voted mostly by the senior subjects with ages between 51 and 63, while lesser relevance was given to this dimension by the subjects with ages between 36 and 40.

To evaluate the measures to be potentially offered, we included another multiple choice question relating all the dimensions that were listed before. Because we intended to make the kinds of measures clearly understood by every subject independently of his literacy, we included only the more standards known ones, such as the average, the minimum, maximum and values and the standard deviation. As usual, we also allow the subjects to leave comments and suggestions about the measures to be included in the visualizations. Fig. 9 shows some of the results obtained.

The average (AVG) was the highly scored measure for all the dimensions followed by the total (count), maximum (Max) and minimum values (Min). For instance, as showed in Fig. 9, for the rating dimension, the average followed by the maximum value and the total (the counting) were the most chosen (with 33, 16 and 18 votes respectively) as independent measures to be made available, although some of the subjects (6) indicated that both the standard deviation (SD) and the average should be informed in the visualization of this dimension. A fewer number of subjects optioned the possibility of the data being presented with more than one measure. Almost for all the dimensions, just one subject tended to indicate a more complex range of measures.

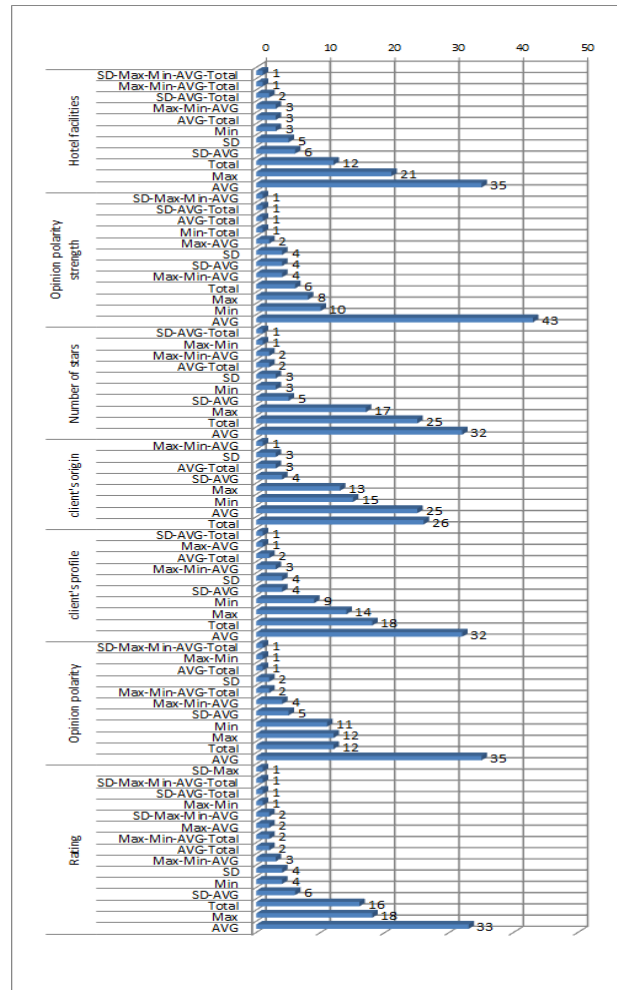


Fig. 9 Measures chosen to dimensions

#### IV. DISCUSSION

In addition to the data analysis, the results obtained through the answers of the survey deserve a more depth reflection.

Almost two third of the respondents always or frequently use Web 2.0 portals for booking accommodations, while 84 percent of the respondents usually read the reviews. Moreover, most respondents (77%) consider decisive the content of the online reviews in booking an accommodation. This result is in line with Gretzel et al. [4], who describe that almost 78 percent of respondents of a survey think other travelers' online reviews are extremely or very important for deciding where to stay.

In the questions 1 up to 8, most women preferred the chart 4, while men optioned mostly for chart 2. These questions were directly related with the identification of relevant visual attributes and thus, visual tendencies. A more recent model of low-level vision has been presented by Huang and Pashler [15]. This theory carefully divides a visual search task into two parts: selection and access. Selection involves choosing a set of objects from a scene. Access determines what properties of the selected objects a viewer can apprehend. Although both



operations are implicitly present in previous theories, they are often described as a whole and not as separate steps. Huang et al. [16] proposed that the visual system is capable of dividing a scene into exactly two parts: selected elements and excluded elements. This is the "Boolean map" that underlies their theory. The visual system can then access certain properties of the selected elements in the map. Based on this, we may conclude that women and men possibly create slightly different vision Boolean maps of the data, thus giving that result.

The strength of the polarity and the opinion polarity were the less voted dimensions. They are resulting from the ontological data processing and thus, they are actually metadata. We consider that possibly one of the causes of them did not have been significantly chosen was a straightforward fact: their meaning was not fully clear and comprehended. It is easy to people misunderstand the meaning in linguistic terms, and maybe, the "polarity" word would not have been clearly perceived. According to Hirst [17], writer-based and reader-based views of text-meaning are reflected by the respective questions "What is the author trying to tell me?" and "What does this text mean to me personally?" Contemporary computational linguistics, however, generally takes neither view.

Another interesting finding was that the rating dimension was voted by all subjects with ages between 51 and 63. This dimension is an ordinal one. Hotel ratings are often used to classify hotels according to their quality. There are a wide variety of rating schemes used by different organizations around the world. Many have a system involving stars, with a greater number of stars indicating greater luxury. We think that this possibly happened because people in this age range are much more experienced and have already concluded that the rating is trustworthy.

In terms of measures applied on dimensions, most of the subjects tended to choose the average as the main one. In this case, maybe this happens because it is a kind of measure that is openly adopted and known. It is learned (how to be calculated) in the first years of school. Moreover, it gives people a rough clue about the data general tendency. In short, it is simple and widely used.

Some of the findings referred in Section Traveler's profile and online booking behavior were fully confirmed in our survey. Although we did not evaluate all the same parameters that Gretzel et al. [4] did, some of ours may be directly correlated to theirs such as client's profile and the opinion date. These dimensions were also significantly indicated as being desired to be visually mapped.

One of the most surprising finding was that users preferred traditional ways of visualizing online reviews of the accommodation sector, such as those offered by tripadvisor.com or booking.com (plain text, basic bar chart and extremely summarized information), instead of more advanced graphical proposals. According to Hullman et al. [18], a user's sense of what is expected from them in using a visualization, including whether to use analytical or intuitive reasoning and the extent to which they should "let the graph

do the work" can be a powerful influencer on how they structure their interaction. Introducing novice visualizations to an end-user maybe induces unproductive levels of frustration, thus not resulting well the influencer premise and making people tend to refuse primarily more advanced data visualization. The fact of the data not being presented in a summarized way can explain this result.

Finally, we also analyzed all the comments received, and we were able to detect some patterns and interesting suggestions:

Use of another color combination with high contrast and avoiding any kind of gradation;

Use of color according international standards;

Horizontal legends instead of vertical ones;

Inclusion of 3D;

Inclusion of more labels with text;

Use of guiding lines in graphs;

Make available the age of the person who opinioned and the total of comments;

Use of glyphs to map data (such as Chernoff [19] faces, for instance).

Finally, we consider that this discussion is relevant to both scholars and practitioners because the graphical representation of online reviews is crucial for helping the decision-making process.

## V. FINDINGS

Based on the results of this exploratory study, we may highlight the most relevant findings. They are categorized according the main topic they focus.

### A. Visual Attributes and Techniques

- Users prefer the information visualization techniques presented in stacked bars and radial graphs to visualize data from online reviews in the accommodation sector. The combination of the relevant perceptual codification for both charts results in the following visual attributes needs to be explored: position, size, shape, color, proximity, connectivity and animation. Thus, data visualization to be implemented should combine both radial graph and bar chart visual techniques, being possibly a kind of hybrid one from these.
- Graphical representation using tree-map technique was not clearly denied to visualize data from online reviews in the accommodation sector: Independent of the visual attribute (e.g. color and shape) under evaluation, users rarely chose the chart generated with the tree-map technique itself. Although the tree-map technique has few adherents, most of its perceptual coding was indirectly chosen through the stacked bars and radial graph. This fact indicates that it is necessary more research to reach some clear conclusion.
- The containment is one visual attribute that is not strongly present in the other two visual techniques, thus, it is not suitable to use in the visualization of this kind of data.
- The same situation happened in the case of the pie charts:

- The rough comparisons between sizes or angles of the slices are not helpful to map the data in this case. Multiplying the number of pie charts magnifies the failure by making it virtually impossible to see how the values changed through concept of ontology.
- Users prefer traditional ways of visualization instead of advanced graphical proposals
- Most of the respondents (70%) prefer simpler and standard ways of visualization (e.g. bar graph with percentage) proposed by the traditional websites than graphics with state-of-the-art information visualization techniques.
- Although this result was reached, considering what was discussed in Section V, and that studies related [17] to learning from visual displays in various subfields of psychology suggest that introducing cognitive difficulties to visualization interaction can improve an user's understanding of important information, it is highly recommended that the resulting visualizations be based not only on standard charts, such as the stacked bar.

#### B. Dimensions and Measures

- The date of the opinion is a crucial dimension to be included in the visualizations, which means that a temporal visualization of the data must be produced (animations, time-slicing, and other temporal visual and interaction techniques should be considered).
- The average, total (counting), maximum and minimum values should be considered as main measures of the dimensions to be made available in the visual representation.
- Some dimensions labels should be changed in the sake of avoiding linguistic or lexical misunderstanding of their meanings. This is especially important when the dimension (e.g. strength of the polarity and opinion polarity) is a metadata.

## VI. FINAL REMARKS

This paper presents the results of a survey about data visualization provided from online reviews in the accommodation sector. The findings evidence the visual model and the cognitive response of users of accommodation booking websites. These findings are useful to design web interfaces according to the real users' preferences.

Although the outcomes delivered by this work give us relevant hints about the visual design to pursuit, it also demonstrates that the process of modeling the data visualization is still far from being a solid path. Even adopting current Information Visualization modeling methodologies and taking into account its best practices, our results must be assumed only as guidelines and does not guarantee completely the success in future visual representations to be implemented subsequently.

This is a pitfall pointed out by the dichotomy present in some subject's answers and comments in the survey, such as: the subjects show a clear preference for charts, however when directly asked and confronted about it, they denied. Maybe

this drawback is due to the significant discrepancy between potential end-users profiles (including cultural) and goals. Anyone can be an accommodation sector data visualization consumer. This imposes a harder degree to identify visual and cognitive demands from these end-users and possibly limitations to the current modeling methodologies.

Finally, along this research we also identify a set of limitations: (a) More than 95 percent of the respondents were Portuguese or Brazilian. Although we did not find remarkable differences between Portuguese and Brazilian users' preferences, it is possible to get different results with people from other cultures. (b) Considering that most of the questions in the survey had an open field to the respondents insert comments (e.g. express sentiments or doubts), some respondents complained that the questionnaire contained complex terms (e.g. polarity and concept of ontology). Such a fact probably can have limited the compression of some questions. (c) Color should be presented in a simple scheme - the magic number [16], for short term memory is seven plus or minus two. When using color in interfaces the number should not exceed five plus or minus two. (d) The number of colors on the screen should be limited accordingly. If the user is overwhelmed or confused by too many colors vying for his attention, he is unlikely to develop an effective mental model.

The next steps of this research include the use of the main findings described in this paper as input in the development of a prototype to provide visual representation of the online reviews. Moreover, we intend to use this prototype to run further tests to reach a finest tuning between the end-user visual model and the data mapping and displaying offered. We also want to know better how the views of the data presented to hotel managers and hotel customers should be both at its visual and interaction levels distinguished.

## APPENDIX

### RESPONDENT PROFILE

Characteristic	Frequency	Percentage
Gender		
Male	73	67
Female	36	33
Age		
18-25	7	6.4
26-30	15	13.8
31-35	26	23.9
36-40	26	23.9
41-45	16	14.7
46-50	8	7.3
51-63	11	10.1
Education		
Doctoral degree	32	29.4
Master's degree	28	25.7
Graduation/ Bachelor degree	32	29.4
High School	17	15.6

## REFERENCES

- [1] Gamon M., Basu S., Belenkov D., Fisher D., Hurst M., and König A. C.. BLEWS: Using blogs to provide context for news articles. In AAAI Conference on Weblogs and Social Media 2008, pages 60–67.
- [2] Draper G. and Riesenfeld R.. Who votes for what? A visual query language for opinion data. IEEE Transactions on Visualization and Computer Graphics, 2008, 14(6):1197–1204.
- [3] Wu Y., Wei F., Liu S., Au N., Cui W., Zhou H., and Qu H. OpinionSeer: Interactive Visualization of Hotel Customer Feedback. IEEE Transactions on Visualization and Computer Graphics, . Nov-Dec 2010, 6, 1109-1118.
- [4] Gretzel U., Yoo K. H. and Purifoy M.. Online Travel Review Study, Role & Impact of Online Travel Reviews. Laboratory for Intelligent Systems in Tourism Texas A&M University, 2007.
- [5] Soukup T. and Davidson I. Visual Data Mining: Techniques and tools for data visualization and mining. Ed. Wiley, ISBN-10 0471149993, May 2002.
- [6] Carvalho E. and Chaves M.. Exploring User-Generated Data Visualization in the Accommodation Sector. 16th International Conference Information Visualisation. Montpellier, France, 10-13 July, 2012, pages 198-203.
- [7] Gamon M., Aue A., Corston-Oliver S., and Ringger E. Pulse: Mining Customer Opinions from Free Text. In: Proceedings of the International Symposium on Intelligent Data Analysis IDA, 2005, Nr. 3646, p. 121–132.
- [8] Lam H., Bertini E., Isenberg, P., Plaisant, C. and Carpendale, S. Seven Guiding Scenarios for Information Visualization Evaluation. TR 2011-992-04, Department of Computer Science, University of Calgary, 2011.
- [9] Ware C.. Information Visualization (3th Edition): Perception for Design (Interactive Technologies). 2012. ISBN: 9780123814654
- [10] Tweedie L. Characterizing Interactive Externalizations”, Proc. Conf. Human Factors in Computing Systems, 1997, pp. 375-382.
- [11] Chi E.H. A Taxonomy of Visualization Techniques using the Data State Reference Model. Proc. IEEE Symposium on Information Visualization, 2000, pp. 69-75.
- [12] Spencer R. Information Visualization: Design for Interaction (2nd Edition). Prentice Hall. 2007. ISBN-13: 978-0132065504.
- [13] Card S. K., Mackinlay, J. D., and Shneiderman, B. Readings in Information Visualization: Using Vision to Think. Morgan Kaufmann Publishers, San Francisco, 1999.
- [14] Chaves M., Freitas L. A. and Vieira R. Hontology: a multilingual ontology for the accommodation sector in the tourism industry. Proceedings of the 4th International Conference on Knowledge Engineering and Ontology Development, Barcelona, Spain, 4-7 October, 2012.
- [15] Huang L. and Pashler H. A boolean map theory of visual attention. Psychological Review 2007, 114, 3, 599–63.
- [16] Huang L., Triesman A. and Pashler H. Characterizing the limits of human visual awareness. Science 2008, 317, 823–825.
- [17] Hirst G. The Future of Text-Meaning in Computational Linguistics. Lecture Notes in Computer Science, Volume 5246/2008, 3-11, 2008.
- [18] Hullman J., Adar E., and Shah P. Benefiting InfoVis with Visual Difficulties. Proceedings of InfoVis'11, Providence, RI, October 23-28, 2011.
- [19] Lee M. D., Reilly R. E., Butavicius M. E. An empirical evaluation of Chernoff faces, star glyphs, and spatial visualizations for binary data. ACM International Conference Proceeding Series; Vol. 142, Proceedings of the Asia-Pacific symposium on Information visualisation - Volume 24, 2003.

concerned with the area of computer graphics with special focus on information and scientific visualization of data.



**Marcirio Silveira Chaves** was born in Caçapava do Sul, Brazil and holds a Ph.D. in Computer Science by the University of Lisbon (2009), a Master degree in Computer Science by Catholic University of Rio Grande do Sul (2004) and a bachelor degree in Informatics - Systems Analysis by University of Vale do Rio dos Sinos (2002).

He is researcher and professor at UniversidadeNove de Julho – UNINOVE, Brazil. He works in the area of Project Management Strategy in the research line of Project Management with a project titled Managing Information Technology 2.0. He has also published papers in the areas of Information Systems, Social and Semantic Web, Software Engineering, Natural Language Processing and Information Visualization.



**Elizabeth Simão Carvalho** was born in Rio de Janeiro, Brazil and holds a PhD in the field of Programming and Informatics Systems Engineering (2008), a MSc degree in the field of Computer Science (1994), both from the University of Minho, Braga, Portugal, a post-graduation in System Analysis (1990) and graduation in Electrical Engineering (1984), both from University Veiga de Almeida,

Rio de Janeiro, Brazil.

She is a scientific researcher at CIAC – Research Center for Arts and Communication of the University of Algarve, Algarve, Portugal and an Assistant Professor at University Aberta, Lisbon, Portugal. Her research is