Systematic Variation of Preattentive Attributes to Highlight Relevant Data in Information Visualization

Original
Systematic Variation of Preattentive Attributes to Highlight Relevant Data in Information Visualization / Barrera-Leon, Luisa; Corno, Fulvio; de Russis, Luigi. - ELETTRONICO. - (2020), pp. 74-79. (Intervento presentato al convegno 24 International Conference Information Visualisation (IV 2020) tenutosi a Melbourne, Australia nel 7-11 Sept. 2020) [10.1109/IV51561.2020.00022].

Availability:
This version is available at: 11583/2842762 since: 2021-03-17T11:01:49Z

Publisher:
IEEE

Published
DOI:10.1109/IV51561.2020.00022

Terms of use:
This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)
Abstract—In information visualization (InfoVis), the Visualizers (graph designers and creators) have to consider multiple parameters, such as colors and graphic symbols, to obtain a chart that correctly represents a data set. Along with this, visualizers must adequately select the combination of these range of parameters to drive the observers’ attention to the relevant data. When a visualizer drives the attention to relevant aspects of the information, she is providing a starting point to read the graph; this focus point might help the observer to complete the task faster and more efficiently, minimizing distraction from unimportant information. Contemporary tools for InfoVis help visualizers to a certain extent, but most of them currently do not provide insights or suggestions about the modifications needed to drive data attention. This article presents the preliminary results of an exploratory approach to draw the attention to some specific data subset selected by the graph creator, through a systematic variation of some preattentive attributes (i.e., color, texture, and orientation). As a first simple method to validate the feasibility of the approach, a set of charts is created from the same source data, with exhaustive variations on preattentive attributes. All generated charts are then automatically evaluated using a salience map algorithm for data analysis images, to identify their focus attention point. After that, the algorithm chooses the chart that best emphasizes the data subset initially specified by the visualizer. To validate our approach, we have implemented a prototype tool, and preliminary results confirm that it is possible to systematically change the attention area of a chart.

Index Terms—saliency, perception, information visualization, preattentive, highlighting

I. INTRODUCTION

According to Rodrigues et al. [1], Information Visualization (InfoVis) “provides faster and user-friendlier mechanisms for data analysis, because the user draws on his/her comprehension immediately as graphical information comes up to his/her vision”. This means the human vision system is strongly related to cognitive and memory processes that take action in data comprehension [2] and interpretation [3]. Even, attention controls what visual information is held and stored [4]. Also, Rodrigues et al. [1] said that carrying out a conscious management of the human vision system, mainly the preattentive process, should be the first stage towards the creation of a chart. The preattentive process takes place at an early stage in the human vision process and defines potential attention areas to be detailed. In InfoVis, the preattentive process is related to the handling of some visual elements called preattentive attributes [5] (e.g., color, size, shape, spatial position).

All this knowledge about preattentive processes has not been thoroughly explored in the InfoVis area, also due to the rapid expansion of InfoVis techniques on different knowledge areas not directly related to design or cognitive processes. However, according to Jänicke et al. [6], "every visualization creator would be curious about how a visualization is perceived and what the observers learn from it". This viewer perception might be measured by techniques such as saliency detection. Visual saliency models could predict which areas attract the observer’s visual attention [6]. With this information about where the observer attention will be focused at first, the creator could know beforehand where his chart will start to be read. Also, she could know how well the chart is aligned with the relevant data. Some contemporary InfoVis tools (e.g., Tableau) use the preattentive process knowledge in an implicit manner. For instance, they present charts with sequential color palettes to highlight, in a natural way, the most important points (e.g., low and high subset) [7]. However, these tools do not provide a mechanism to know if the selected chart (e.g., subset position, shape, color palette) is the one that better emphasizes the relevant data or how is it perceived by the observer.

Our main goal is to explore possible approaches for helping graph creators and designers (hereinafter, visualizers) to draw the observer’s attention to specific relevant data, selected by them at design time. To assess the feasibility of the approach, this paper proposes an exploratory experiment, in which we first perform systematic chart variations, and then evaluate their effect using a saliency map algorithm. For the sake of this first experiment, we consider bar charts, only, and consider three preattentive attributes, namely, color, spatial position, and orientation. Chart variations are evaluated using the saliency map algorithm proposed by Matzen et al. [8] who modified the standard Itti method [9] to adapt it to the particular characteristics of InfoVis images. The chart that better emphasizes the data subset specified by the visualizer is finally selected. With this approach, we expect to obtain
insights about the feasibility of moving the attention area in a chart, and about which attributes could generate a greater impact towards this.

To get further insights about this exploratory approach, we implemented a prototype tool. In the tool, the visualizer selects the data subset that she wants the observer to focus on. Then, the tool generates several chart variations, keeping the same chart semantics but modifying: a) color palettes, using sequential and discontinuous palettes; b) spatial position, changing data subsets in the X-axis; and c) the orientation, by swapping X and Y axis. All chart variations are evaluated with a saliency map algorithm [8], to allow the tool to choose the chart with the largest salience area in the originally selected data subset.

II. BACKGROUND

In this section, we present an overview about the preattentive process in information visualization and some fundamentals about salience maps, which are the basis for our exploratory approach.

A. Preattentive Process in InfoVis

According to Healy et al. [3], one of the goals of visualization is to support visual analysis, exploration, and to discover novel insights. To achieve this goal, it is important to consider human perception. How humans perceive images and charts details can impact a viewer’s efficiency and effectiveness [3] of their cognitive process. Human vision system attention has various mechanisms that determine which regions of an image are selected for more detailed analysis. One of this mechanisms is the preattentive process that occurs in the first 200 milliseconds of the vision process. The preattentive process computes each preattentive attribute in parallel and then combines them to select specific regions that are perceived without any conscious effort [9]. According to some researchers [4], [5], the classification of preattentive attributes are: Form, which bundles line, orientation, length, width, collinearity, size, curvature, and others; Color, including hue and intensity; Motion including flicker and direction; Spatial position made up of 2D position, (stereoscopic) depth, depth or convex/concave shape from shading.

By knowing how the preattentive process works, it is possible to change the preattentive attributes for driving the viewer attention to a specific point. Rodrigues et al. [10] support this idea by establishing those preattentive stimuli that constitute the first step in the vision expression process, and by exploring how this attentive selection and vision support charts cognition.

Some of the research relating preattentive processes with InfoVis focuses on how to handle preattentive attributes to: highlight a subset in a specific chart type [11], highlight a semantic link across graphs [12], or to increase the amount of information present in one chart [13]. Most of the preattentive attributes used to highlight are Color and Form (e.g., size, shape, orientation). In Carenini et al. [14], the authors proved that highlighting some data on a chart improves visualization. They use preattentive attributes (emphasis, bolding, shapes, etc.) to guide the viewer to focus on specific data members. This guided focus improves viewer’s performance in a chart analysis test.

On the other hand, some researchers use algorithms based on the human preattentive process to predict and measure where the focus attention area is located in specific chart types. To do this, they mainly use two preattentive computational models: saliency maps to predict, and eye-tracking to measure. For instance, Raschke et al. [15] used eye tracking to analyze perceptual and cognitive processes of visual tasks on bar charts. In particular, the authors use eye-tracking techniques to detect areas of interest. Lastly, Alberts et al. [16] designed an experiment to study how certain design elements affect where people look on dashboards. They found some insights about how preattentive attributes influence dashboards comprehension.

The closest work to ours is presented by Janicke et al. [6]. They used the model proposed by Itti et al. [9] to evaluate the quality of visualization images, which allows the visualizer to choose the best visualization from a set of alternatives. In this research, the user creates a relevance mask determining the most significant points in the chart. After this, the user can also define the degree of influence on three image component channels (color, intensity, and orientation). Then, Janicke et al. matched the Itti saliency map with this user relevance map to know how well visual salience matches data relevance. However, this approach helped to select the most focused visualization alternative, among the ones produced by the visualizer. We are instead aiming at adapting the chart to the the visualizer relevance map using systematic changes.

These approaches focuses on the creation of patterns, models, or design guidelines so that a visualizer is informed about the best way to design her chart type or data. In comparison, our work is focused on supporting visualizers to “adapt” their charts, by using chart variations rules as preattentive attributes, evaluation mechanisms such as salience maps, and various techniques to select the closest option to meet visualizers’ requirements in relation to highlighting relevant data.

B. Saliency

Koch and Ullman [17] proposed a biological model to explain the first step in the human vision process, specifically the preattentive process (section II-A), describing how the human vision focuses attention across a visual scene. In this context, Saliency represents the feature map that determines which locations on an image are more prominent from their surroundings [2]. To predict this saliency on an image, Koch and Ullman identified how some scene attributes such as color, spatial position, and direction among others, can be represented in early selective visual attention maps. The combination of these maps, called saliency map, “combines the information of individual maps into a global measure of conspicuity” [17]. Itti [9] created a computational model for detecting saliency based on the Koch and Ullman approach.
Both models, biological and computational, have been widely used in many algorithms [18]. However, these algorithms were created for natural scenes images, and such images’ properties are quite different from those of InfoVis charts [8]. In fact, some research aimed at modifying such a saliency model to account for the characteristics of InfoVis images. One of them is presented by Bylinskii et al. [19], that presented two automated models to predict the relative importance of the elements (e.g., title text, axis text, data points) in data visualization and charts designs. The models are neural networks trained with results on “human clicks” and with the “importance annotations” tests. In “human clicks” tests, a blurry image is shown to a participant, and then she clicks on different parts of the image to reveal small regions. In the case of “importance annotations” tests, participants were asked to label relevant regions in a design, based on Graphic Design Importance (GDI) dataset [20]. With these two inputs, their prediction models learn to localize the titles and correctly weight the relative importance of different design elements [19]. To sum up, the two trained models can predict saliency based on the perceived relative weighting of design elements, disregarding the human perception biological process.

Another study about saliency maps for InfoVis has been developed by Matzen et al. [8]. They proposed a Data Visualization Saliency Model (DVS) using a modify Itti model. The Itti algorithm uses the RGB model to compute the saliency map, but to better approximate human perception, DVS transforms the input image into the CIE LAB color space. Also, they combined the Itti-modified algorithm with a Text Saliency Model, because the observers in InfoVis devote a disproportional amount of attention into regions containing text. DVS was tested with the MASSVIS dataset [21], with CAT2000 [22] and with some images from its creators. Additionally, the study of Bylinskii [19] is supported by neural network training with validations performed by crowdsourcing, without having a controlled testing environment. Furthermore, Matzen algorithm combines Bylinskii’s findings regarding the high observer’s attention on the text regions in InfoVis images. In our approach, we chose to adopt this latter algorithm for evaluating the salience of the generated charts, as an evaluation tool for predicting image salience, and therefore observers’ focus.

III. APPROACH DESCRIPTION

Our approach starts from a data-set to be visualized, and one data item that the visualizer wishes to highlight through changing preattentive attributes. The objective is to identify a chart, whose attention is focused on the data of interest, by performing a systematic manipulation of selected preattentive attributes (color, texture, and orientation). Our exploratory approach starts with three inputs: the dataset, the focus data subset, and the color palettes profiles. The former constitutes the data that is going to be visualized. The latter is a subset of the dataset defined by the chart creator defines, whose information is to be highlighted. The visualizer may choose different color palettes, that will be individually processed. The output is a group of charts, one for each color palette, that best emphasize the data subset specified by the visualizer.

![Fig. 1. Process of preattentive driven InfoVis](image)

The proposed approach is composed of five main phases, as illustrated in Fig. 1: 1) Load initial data, 2) Charts Systematic Construction, 3) Saliency Maps Generation, 4) Salience Rating, and 5) Plot Final Charts.

In this preliminary investigation of the proposed approach, we selected three preattentive attributes to change in static charts: color, texture, and orientation. We choose color because, according to Triesman and Gormican [23] and Ware [4], to highlight a specific target, it is more natural to use color variations than other preattentive properties. Also, color is the most dominant factor when computing a saliency map [3]. We decided to use texture because, according to Healey [3], it’s the second most dominant factor in a saliency map. Additionally, some preliminary tests were performed, and when a texture is applied, it is possible to reach a focus area when the chart color palette does not produce a saliency focal area. Finally, with orientation, part of form preattentive classification (Section II), we are interested in taking advantage of the chart ability to change its orientation without changing its meaning, and test if it changes or improves the salience in a certain way.

In this initial experiment we also limited our attention to one type of charts, namely bar charts.

A. Load initial data

Initially, our approach requires three inputs: the data set to be plotted, the specific data to be highlighted and the profiles of the palettes. About the data to be focused, in this initial approximation, it should be a categorical data selected by the visualizer.

Regarding the palette profiles, the visualizer could select one through pre-set color palettes. As in most visualization tools, these palettes can be of various types, among which are sequential, divergent and qualitative. We limited the research using only qualitative palettes taken from COLOURlovers [24] and filtered by those who were related to the keyword “data”. We selected categorical palettes because the positions of the colors can be shifted without affecting the meaning of the colors, which would happen with a sequential palette whose colors must be “ordered”. For each palette, a profile was constructed to determine which color in the palette has the most salience, and if the salience increases using textures or changing the position on the chart (see Fig. 2). To create such profiles, the Matzen salience algorithm [8] was used. By using a bar chart as a baseline, the selected palettes were applied and then passed through the salience algorithm. At the end of this...
In this step, the colors of each palette were ranked for prominence. We repeated the same process with texture and orientation attributes. All these values can be pre-computed are are not dependent on the specific data-set.

Table I presents an example of a palette profile. ‘Dominant color’ represents the color position into the palette of the most salient color. The other properties are marked with a one (1) if their presence affects salience in the specified palette. For example, the Data Fact palette (see Table I) has a clean saliency with vertical and horizontal orientation, but without texture. Clean saliency means that only has one dominant color, on a specific position (vertical or/and horizontal), and with or without texture. In Fig. 2, the dominant color is number four (4) in the palette, on the vertical (Fig. 2.a) and horizontal (Fig. 2.b) position, both without texture. For the prototype, we didn’t use the palettes without a clean saliency color in one or more of the combinations.

Table I

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dominant color</td>
<td>4</td>
</tr>
<tr>
<td>Vertical</td>
<td>1</td>
</tr>
<tr>
<td>Horizontal</td>
<td>1</td>
</tr>
<tr>
<td>Vertical with texture</td>
<td>0</td>
</tr>
<tr>
<td>Horizontal with texture</td>
<td>0</td>
</tr>
</tbody>
</table>

B. Charts Systematic Construction

In this step, the goal is to systematically create charts, based on the visualizer’s requirements (color palette and data to be highlighted). About the colors, the visualizer selects a set of possible palettes to use in the chart. By taking the color palette profiles as input, a systematic chart creation is carried out. At every step, the chart is created by using the color with the highest salience for representing the data to be highlighted. Then, the remaining colors in the palette are shifted through the other data subsets. Several candidate charts are created, taking into account whether the palette profile establishes that salience increases with texture and/or orientation.

C. Saliency Maps Generation

In this step, we generate salience maps for each chart variation created in the previous stage. To accomplish this, as reported on the study on salience map models and algorithms (see Section II), the selected algorithm was DVS [8], implemented in Matlab language. The algorithm output is the image of the chart, enriched by a salience map layer.

D. Saliency Rating

In this step, each salience map is evaluated to identify which charts have the highest saliency values on the area corresponding to the data subset to highlight. The DVS algorithm has a function that supports this ‘proximity’ evaluation. This function compares a given set of coordinates with the same coordinates on the salience map. The coordinates of the data subset to be highlighted are sent to this function; the function returns a value between 0.0 and 1.0 that measures the weighted overlap of the data subset coordinates and the chart salience areas. The final result of this step is a list with the percentage of proximity value per each chart variation.

E. Plot Final Charts

The last step is the selection of the best chart. For each palette selected by the visualizer, the chart with the best proximity percentage is chosen (see Section III-D). The winning charts are shown to the visualizer, and she can choose the preferred one.

IV. RESULTS

To validate our approach, we implemented a prototype tool in Matlab using common libraries for data analysis and visualization (e.g., App Designer, Plots) and we generated salience maps with the DVS library [8]. The goal here is to understand the behavior of the preattentive attributes and the feasibility of moving the attention area starting from a baseline chart. In this first round of experiments, we use bar charts as a baseline on which to apply the systematic variations.

To carry out our evaluation, we selected the Coronavirus (COVID-19) dataset [25], which consists of data from COVID-19 Deaths Worldwide as of May 29, 2020. We plotted the five countries with the most casualties (USA, UK, Italy, France, and Spain) vs. the number of deaths (103,330, 37,837, 33,142, 28,662, and 27,119 respectively). This combination was selected because it generates a graph with a considerable variation between subsets. The other countries have a difference, in the number of deaths, of at least 17,000 with respect to the selected five, which makes their bars too tiny compared with the others. As the bars are small, in the chart they saw as a line and not as a bar, the changes in color and texture could be difficult to perceive.

Fig. 3 shows the prototype interface. The visualizer must first select the database for which she wants to plot the chart. For the prototype, only Coronavirus Deaths were tested. Based on the selection of the dataset, the visualizer must then select the data subset to be highlighted. The next step is to choose some possible color palettes (three, in the example) that the visualizer wants to use on his chart (see section III-A). After that, the visualizer can execute the algorithm presented in section III. For each color palette selected by the visualizer, a minimum of one and a maximum of 8 charts are generated. Four of these charts correspond to the states...
in which the palette obtains salience in the dominant color (see section III-B). The other four correspond to variations on the remaining colors (not predominant) of the palette. Fig. 4 shows an example of those variations, with the same palette profile presented in Table I, and Italy as a the data selected to be highlight (third data in the chart). Chart b in Fig. 4 is the chart with the best ranking. The final result is composed of the charts selected by the algorithm, one per palette, whose salience map had the highest percentage of proximity to the data selected by the user (Fig. 3). Additionally, the interface displays the salience maps of each of those charts.

Based on the prototype results, we found some insights about preattentive attributes impact. Regarding to preattentive attribute Color, although we only worked with palettes that had a predominant color, we found that this predominance could be stronger or weaker depending on the neighboring colors. In Fig. 4, it can be seen how the salience is more precise in the color combination b and less intense in d. It is common, due to lack of evaluation tools, that if the viewer wants to highlight some data, she uses the “strongest” color. Nevertheless, it is possible to have better accuracy, also knowing the incidence of the neighboring colors.

Concerning to preattentive attribute Texture, in most of the revised color palettes, the texture made the salience region wider. Also, in some palettes, only it is possible to obtain a salience focus on one data if that one has a texture. Otherwise, the palette has only a soft salience in the text on the charts. In the “data Color Pie” color palette chart on Fig. 3) the palette has a texture because without a texture, the saliency would be on the title, but with the texture is on the data. Nevertheless, also the texture increase the salience area to much. Even so, the salience is not as clean as with the other graphics, because the saliency region is on three data subsets. A future step could be to achieve systematic handling of the textures to achieve a more precise salience.

With respect to the preattentive attribute Orientation, we found that the salience is consistent, with the same predominant color, when the orientation is changed. However, since the Matzen et al. algorithm has a text recognition component when some of the data has a long name, the salience region is enlarged. Fig. 4.c has the same color order than b, but the saliency area is more expanded at the bottom. That happens because the name of the four (4) data is large, and it creates a black text area which will be highlighted most prominently by the Matzen algorithm.

We found that it is feasible to move the attention area to a specific data subset. The visualizer can select the data to be highlighted and the color palettes of her preference. In the end, she gets information, in the design stage, about what could be the impact of its color selection on the chart perception. That is why palette profiling helps the visualizer to evaluate more than one alternative in the positioning of the colors in the chart automatically. This support is not available in the current visualization tools.

Unfortunately, due to the current pandemic situation, we could not test our approach with final users. Since salience algorithms are a model of human visual perception, for the time being we are relying on salience maps as a proxy of user attention. However, most of the work on these predictive algorithms in the InfoVis area is relatively recent (3-4 years), and it has not been extensively studied. Therefore, the next step in our work will be to evaluate the prototype with end-users. The planned user studies will evaluate both the precision of the chosen salience map algorithms and the overall feasibility and effectiveness of the proposed approach.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented preliminary results of an exploratory approach about how to draw the attention to a specific data subset selected by the visualizer, by means of a systematic variation of preattentive attributes. To accomplish our goal, we created a set of chart variations using preattentive attributes (color, texture and orientation). Then, each graph
was evaluated with a salience map algorithm created for InfoVis images, to know which data subset was attracting the observer’s focus. Finally, we selected the one exhibited a focus closest to the data subset selected by the graph creator. In order to validate this approach, we developed a tool in Matlab and used it to test a Coronavirus (COVID-19) dataset. The results show that it is possible to drive the viewer to focus her attention on a relevant information selected by the visualizer. The tool can generate diverse charts depending on which relevant data was selected to be focused (e.g. Italy or USA), and the visualizer preferred color palettes. Concerning the preattentive attributes impact on the chart attention area, according to the results, color was the attribute that had a bigger impact on highlighting the proper data subset. Another insight was that the salience is consistent also when it is changed the chart orientation.

In conclusion, we demonstrated that it is possible to automatically modify certain aspects of a graph to meet some specific needs of its creator. To fully prove our approach, it will be necessary to do a user evaluation and to contrast the results obtained in this first investigation. Therefore, it is necessary continue to looking for a salience algorithm that are better suited to the characteristics of InfoVis images and obtain a better precision on which data is being highlighted.

REFERENCES