Incidental Visualizations: Pre-Attentive Primitive Visual Tasks

João Moreira
joaopedro.moreira@inesc-id.pt
INESC-ID Lisboa, Instituto Superior Técnico, Universidade de Lisboa

Daniel Mendes
danielmendes@tecnico.ulisboa.pt
INESC-ID Lisboa, Instituto Superior Técnico, Universidade de Lisboa

Daniel Gonçalves
daniel.goncalves@inesc-id.pt
INESC-ID Lisboa, Instituto Superior Técnico, Universidade de Lisboa

ABSTRACT

In InfoVis design, visualizations make use of pre-attentive features to highlight visual artifacts and guide users' perception into relevant information during primitive visual tasks. These are supported by visual marks such as dots, lines, and areas. However, research assumes our pre-attentive processing only allows us to detect specific features in charts. We argue that a visualization can be completely perceived pre-attentively and still convey relevant information. In this work, by combining cognitive perception and psychophysics, we executed a user study with six primitive visual tasks to verify if they could be performed pre-attentively. The tasks were to find: horizontal and vertical positions, length and slope of lines, size of areas, and color luminance intensity. Users were presented with very simple visualizations, with one encoded value at a time, allowing us to assess the accuracy and response time. Our results showed that horizontal position identification is the most accurate and fastest task to do, and the color luminance intensity identification task is the worst. We believe our study is the first step into a fresh field called Incidental Visualizations, where visualizations are meant to be seen at-a-glance, and with little effort.

CCS CONCEPTS

• **Human-centered computing → Information visualization**; **Visualization theory, concepts and paradigms**; **Empirical studies in visualization**.

KEYWORDS

incidental visualizations, pre-attentive, primitive visual tasks, user study, cognitive perception, psychophysics

ACM Reference Format:

João Moreira, Daniel Mendes, and Daniel Gonçalves. 2020. Incidental Visualizations: Pre-Attentive Primitive Visual Tasks. In *International Conference on Advanced Visual Interfaces (AVI '20), September 28-October 2, 2020, Salerno, Italy*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3399715.3399841

1 INTRODUCTION

With the growth of data surrounding us daily, Information Visualization (InfoVis) may play a crucial role in allowing us to understand relevant information in a timely fashion. The massification of information appliances coexisting with us raises the need for visualizations that allow users to understand information at a glance, during their daily activities. One way to make specific elements of visualizations easily noticeable is by manipulating pre-attentive features, which can be seen without conscious attention. These features (called channels) apply to marks (dots, lines, or areas) , and they vary depending on the chosen mark. By making use of these features, users can spot features at-a-glance, for example, one red circle next to many blue circles will be seen pre-attentively because it will pop out from its neighbors. However, visualization designers assume our pre-attentive processing cannot be used to perceive both these features and information.

Most visualizations make use of users' conscious focus, whether because they require them to interact or because the information is too complex. When visualization displays are not the primary focus, the topic shifts to peripheral displays (which originated in ubiquitous computing). These can show visualizations, but sometimes a peripheral display is just an LED, since their goal is just to convey information through the users' sight periphery. Then, ambient displays came to be (inheriting from peripheral displays), to display information related to its physical surroundings. Finally, ambient information displays (inheriting from ambient displays) were created, to display visualizations that continuously send environment information to users. However, they are still designed to be seen attentively, even though they demand less focus on users.

We take the first step towards an emerging design: Incidental Visualizations. Unlike traditional visualizations, these will be displayed temporarily, designed to be seen at a glance, and without stealing the users' primary focus. We argue that these visualizations are the next step towards seamless cooperation between users and information visualization, resulting in augmented perception. For instance, let us suppose we are at our house, and we are leaving the kitchen. Before exiting, we flip the light switch to turn off the lights. As soon as we starting flipping the switch, an incidental visualization is showed to convey information on the kitchen's energy consumption while comparing it with the rest of the house. If seen at a glance, it won't hinder the user. If all is well, it can be mostly ignored. In the case when relevant new information is displayed, it can be used for decision-making.

The rest of the document starts with a summary of the most relevant topics: InfoVis (the current best practices at visualization design), Cognitive Perception (different techniques to measure our perception), and Psychophysic (visual stimulus, and its effect on humans). We end it with a brief discussion of how all topics come together to end up as a baseline for our work. Then, we present our user study, where we measured how accurately users can perceive information pre-attentively, and how fast they react after each perception. We tested five channels: position (horizontally and vertically, using the dots), length (lines), tilt (lines), size (area), and

AVI '20, September 28-October 2, 2020, Salerno, Italy

ACM ISBN 978-1-4503-7535-1/20/09…$15.00
https://doi.org/10.1145/3399715.3399841

luminance (area). The key contributions of this paper are: which primitive visual tasks allow higher accuracies and lower response times, and a set of design guidelines for future visualizations to be seen pre-attentively.

2 RELATED WORK

Before moving into our experimental design, we present several explicit approaches to test visualizations' utility, which usually depends on the research community [33]. These usually coexist alone, but it is possible to combine different techniques [23, 40]. Then, we will explain the current best practices to create visualizations.

2.1 Information Visualization

In the InfoVis community, to measure visualizations' effectiveness users are requested to execute several search tasks, which can be the identification, or manipulation, of several visual artifacts. Search tasks are easier if they make use of pre-attentive features such as shape, angle, size, and texture. Searching for these features takes almost constant reaction time, regardless of the total number of visual artifacts [31]. However, when several pre-attentive features are placed simultaneously, the results get worse, because they interfere with each other. The only well-known exception is hue and brightness combined [2]. At first, these studies were made without data. Eventually, pre-attentive features were applied in visualizations to see if they could help users explore several regions of data [15, 16].

Studies on information visualization deeply explore pre-attentive graph perception, to convey information quickly. There are four well-known models: feature integration theory, texton theory, similarity theory, and guided search theory. Anne Treisman was one of the first researchers to work on pre-attentive processing. She worked in feature integration theory, which is the most simple to understand. It relates to how we can find and/or manipulate several visual artifacts. Texton theory focuses on the statistical analysis of texture patterns, which is more high-level than Treisman's theory. Then, similarity theory [26] states that search time does not depend on pre-attentive features, but at how easy it is to find one target from its distracters, and the number of information item required. Finally, guided search theory was introduced by [39], who hypothesized that we look at a visualization both top-down and bottom-up analysis, to create different layers of every feature.

2.2 Perception and Psychophysics

In the perception community, instead of analyzing pre-attentive features, studies tackle the overall performance of different visualization by looking at how much attention they demand from users. There are two ways of doing that: formally, and informally [6]. Using the informal approach, users just need to find the differences between several instances of the same chart, since only the data changes. Using the formal approach, several metrics are chosen to measure the effectiveness of visualizations, while users execute visual tasks on them. Accuracy is the simplest metric to assess a graph's utility. In early studies, accuracy was measured by looking at how users could read information from charts, to justify their use against tables [8, 9, 12]. Later, studies focused on measuring users' accuracy at estimating values, or answer semantic questions [1, 4, 11, 22, 28]. The third approach to measure accuracy is by asking users to detect differences between charts, with no changes in data [24]. Following accuracy, the next main metric to assess a chart's utility is response time. While accuracy focuses on how well users understand the information in a chart, response time focuses on how fast users react to that information. Cognitive load is usually associated with the response time. The longer it takes for users to react to a visualization, the higher the cognitive load it conveys. Both accuracy and response time can be joined to measure a chart's effectiveness [3, 18].

In the psychophysics community, studies to assess data visualization focus on measuring if users can detect a particular stimulus and its effect size. The most common methods are the method of constant stimuli and the method of adjustment. In the latter, users are asked to change the stimulus until they stop receiving it [17]. In the former method, stimulus changes randomly to reduce continuity effects [34], allowing users to receive each one unbiased. Although most statistical graphs came from psychophysics and cognitive psychology [19, 27, 30], these metrics are not usually found in the InfoVis community, showing us a gap between psychophysics and InfoVis.

2.3 Non-Quantified Metrics

Besides quantified metrics, there are two methods that can help assess a chart's utility: thinking aloud, and eye-tracking. Eye-tracking gives an insight into how users look at visualizations, particularly at the main spatial location of visual focus, which is difficult for them to communicate. Additionally, it helps to measure how much time they gaze at a visualization [13, 14, 20]. Although eye-tracking alone does not help understand how users perceive visualizations, it can be used alongside accuracy and response time.

Thinking aloud can help to understand the cognitive process as users perceive charts. We measure this metric by asking users to talk as their thoughts flow. This is useful to quantity users' insight and reasoning during visual tasks' execution [21].

2.4 Good Practices for Chart Design

Our human visual system quickly processes high amounts of information. In graph design, we make use of this system by applying gestalt principles of visual perception: proximity, similarity, common region, common fate, continuity, and closure [37, 38]. Every principle tackles different ways our perception has to relate to cognitive artifacts. Proximity happens when we put several cognitive artifacts close together. Similarity happens when cognitive artifacts are similar. The common region happens when several cognitive artifacts are inside well-defined boundaries. Common fate happens when several cognitive artifacts are changing the same way. Continuity happens when several cognitive artifacts form a continuous line. Finally, closure happens when we recognize several cognitive artifacts from a complex image. Every principle allows users to relate information, if needed, in short periods.

Another good practice is to free humans' working memory. It is limited, so it limits cognitive perception. In chart design, it is advised to reduce the number of categorical insights a visualization imposes because it becomes impossible to map many categories while trying to understand the information conveyed. Additionally, the way humans' memory stores information differs from the chart's representation. When users see a representation, what stays longer in memory is not the details but the overall semantics [25]. This phenomenon is called change blindness.

From psychophysics research [4, 6, 19, 24], a rank of primitive visual tasks was created for accuracy and quantitative evaluations. From most accurate, to least: position, length, direction, angle, slope, area, volume, density, curvature, shading, color luminance, color hue, and shape. Since not everything can be encoded using just one primitive visual task, the rule is trying to use the most accurate ones. If information can be encoded using a more accurate task, then it is preferable for the overall visualization.

Last, is to create an aesthetic chart. Aesthetics has proved to significantly affect how users read data [35]. This contradicts Tufte's guidelines, which suggest the removal of everything that is not conveying relevant information [32]. For example, redundant information may be conveyed, or aesthetic visual artifacts may be placed, if they trigger gestalt principles to help relate information semantically.

2.5 Discussion

As seen, there are already well-defined strategies to measure a chart's utility, and several guidelines to make them as useful as possible. Regarding testing methods, we just tackled explicit testing methods; these assess users' perception of quality based on their cognitive processes. Implicit testing methods ask users what is their overall semantic understanding of a chart. Users are required high-level insights into what they perceive, which is important in the endgame scenario where we want to implement incidental visualizations. But, we are still taking the first steps at this topic, which is why we did not approach them in our study.

Different communities try to assess chart utility with their own techniques, but it is always related to accuracy. In InfoVis, accuracy is measured in search tasks; how well can users find targets shown in several charts? In perception, accuracy is measured by comparing charts with their corresponding tables, to see if they ease information comprehension. In psychophysics, accuracy is measured by analyzing how users react to different stimulus intensities. Although each community is different, they try to discover ways to improve users' ability to perceive information, as fast as possible. But there is a gap in research; each community assumes information is always seen attentively. Their research wants to improve how our perception is guided to find information as quickly as possible. However, we believe it possible to understand information pre-attentively, we just do not know how much.

Besides validating charts, there are well-defined guidelines to create them. Applying gestalt principles helps users to quickly relate information and freeing our working memory free added cognitive load during perception. Then, following chart utility studies, there is a well-defined accuracy rank of different primitive visual tasks. Finally, it is important to make a visualization aesthetic, even though a few researchers are against cognitive artifacts that do not convey useful information. Again, these guidelines also assume information is seen attentively. The accuracy rank, for example, was created measuring accuracy and response time without limiting the time every cognitive artifact was available. With our study, we contribute with new guidelines to create charts, so that primitive visual tasks can be executed pre-attentively.

3 USER STUDY

We conducted a user study to test how primitive visual tasks can be executed pre-attentively, that is, without conscious focus. This quantitative value has been debated, but it was defined recently as less than 500ms [36]. In our case, the threshold was 100ms, to make sure users perceived every visualization within their pre-attentive phase. We tested six primitive visual tasks supported by three well-known marks, and varied their channels: dots (horizontal and vertical positions), lines (length and tilt), and areas (size and color luminance).

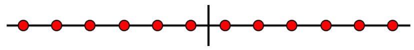
3.1 Method

To gather our data, we designed our user study to be conducted online, through a custom web page, to to make it easier for people to participate. It allowed participants to execute six tests, each for one primitive visual task. The tests were designed to be independent of each other, to avoid data loss in case some fault happened. In particular, each participant's input was registered individually.

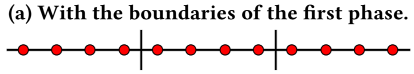
The chosen metrics came from attention mediated testing: accuracy and response time. Accuracy was measured to calculate how far each user's guess was from each encoded value. It was the only metric showing us if users could execute these primitive visual tasks pre-attentively because each encoded value was shown during 100ms. Response time was measured to understand the cognitive load induced by each perception. Then, we used the method of constant stimuli from psychophysics to create several distinct execution orders.

To measure accuracy in worse case scenarios, instead of asking users to estimate values, we asked them to guess in which range each value was presented. We varied the number of ranges by creating three distinct phases: two in the first phase, three in the second, and four in the third. We called these ranges categories, by naming them, thus facilitating users' comprehension. For example, in the horizontal position test, the categories for the first phase were: left, and right. The encoded values were not random since we wanted to use the Latin Square algorithm to generate several orders. However, we chose them so they never ended up in borders between categories. The actual line size depended on the users' displays and window sizes. For the sake of this example, let's assume a 400 pixel-wide line. In the third phase, having four categories (ranges) means that the total space is divided into four bins. The first includes the values between 0 and 99, the second from 100 to 199, and so on. So, the chosen values must never be 100, 200, or 300. What we did was divide each bin into three sub-bins, and we chose the middle points between borders as values. Overall, we end up with 12 points, three for each category (3*4=12). These values are maintained between phases, so we can apply credible statistics tests with the final data. The points are represented in figure 1, and the same logic applies for every other primitive visual task.

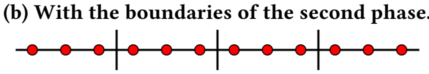
Response time was measured by pressing a keyboard key. At each test's phase beginning, the web page explained its procedure, including which keys corresponded to which categories. Keys were



(a) With the boundaries of the first phase.



(b) With the boundaries of the second phase.



(c) With the boundaries of the third phase.

Figure 1: All points shown for the horizontal position test.

chosen to match each primitive visual task. For example, in the horizontal position test, the key "1" represents the category "Left", which was left to the key "2" that represented the category "Right", during the first phase.

To use the method of constant stimuli, we had to create different orderings for each phase, visual task, and user. We used the Latin Square algorithm to generate a matrix where each row represented a specific order. Since we encoded 12 values in each phase, each row had 12 values. Rows for each user were chosen randomly because the tests were executed online. This matrix was used across each phase in each primitive visual task because the values were always the same (only the visual encoding changes).

Finally, in each task, we measured each answer as if users' perception was linear, which is not. For example, the way we perceive color luminance is different from the way we percieve position. However, this ensured we tested each task equally.

3.2 Primitive Visual Tasks

Before starting each test, users were presented with instructions. The logic was always the same: three phases, each presenting 12 values, one at a time. If users remained confused after reading the instructions, they could choose the sandbox mode, which was implemented to allow users to try each test, to decrease error occurrence in real testing. Besides, no values were registered in this mode.

Horizontal Position. The primitive visual task in this test was to find a point's horizontal position. Each dot was displayed in red color on a horizontal line (Figure 2). Its position varied along the line across 12 locations. In the first phase, users could choose between left, and right. Then, in the second phase, they could choose between left, middle, and right. Finally, in the third phase, they could choose between left, middle left, middle right, and right.

Vertical Position. The primitive visual task in this test was to find a point's vertical position. Each dot was displayed in red color on a vertical line (same as the horizontal encoding, but vertical). Its position varied along the line by 12 locations. In the first phase, users could choose between top and bottom. Then, in the second phase, users could choose between top, middle, and bottom. Finally, in the third phase, users could choose between top, middle top, middle bottom, and bottom.

Figure 2: Example of an encoded value for the horizontal primitive visual task.

Figure 3: Example of an encoded value for the length primitive visual task.

Line Length. The primitive visual task in this test was to find a line's length. Each line was displayed in red color and had one of 12 pre-chosen lengths. The center of the line always originated from the same position, thus avoiding confusing users (Figure 3). In the first phase, users could choose between the categories large and small. It was necessary to explain what large meant. If the line occupied more than half of the total space, then it was large; otherwise, it was small. The total space borders were shown using two vertical black lines, one on each side. In the second phase, users could choose between the categories large, medium, and small. In the third phase, users could choose between the categories large, medium-large, medium-small, and small. With three categories, the total space was divided into three equal parts (four equal parts with four categories).

Line Tilt. The primitive visual task in this test was to find a line's tilt. Each line was red with one of 12 tilts (Figure 4). The angle of each line ranged from zero to 90 degrees. The x and y axes were rendered as straight lines. In the first phase, users could choose between the options: left and bottom. The categories were related to the slope of the line, whether it was closer to the y-axis or the x-axis, respectively. In the second phase, users could choose between left, middle, and bottom. In the third phase, users could choose between left, middle left, middle right, and bottom.

Area Size. The primitive visual task in this test was to find a square's area size. Each square was presented in red as one of 12 area sizes (Figure 5). In the first phase, users could choose between the categories large, and small. As with line length tasks, we had to explain each category. The area is large if it occupies more than half available total space, which is shown as a transparent square with a black outline around the red square. In the second phase, users could choose between the categories large, medium, and small. In

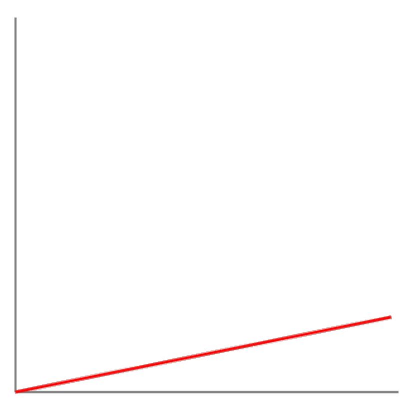


Figure 4: Example of an encoded value for the tilt primitive visual task.

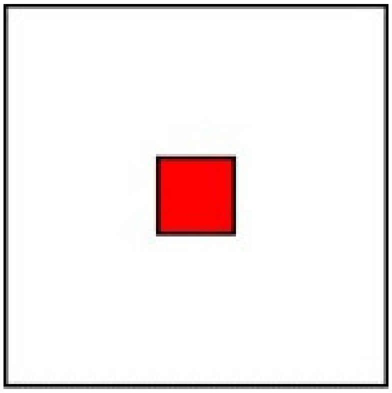


Figure 5: Example of an encoded value for the size primitive visual task.

the third phase, users could choose between the categories large, medium-large, medium-small, and small.

Color Luminance. The primitive task in this test was to find a square's red hue luminance. Each square had one of 12 different luminance values (Figure 6). In the first phase, users could choose between the categories high, and low. Luminance was high if it was in the left half of the color spectrum. The approach for the other categories is the same as with line length and square area. In the second phase, users could choose between the categories high, medium, and low. In the third phase, users could choose between the categories high, medium-high, medium-low, and low.

3.3 Prototype

We created a web page to deploy our user study. The first thing users saw when they entered the web page was a short presentation of the study and its purpose. The web page was created using HTML, JavaScript, and D3.js (https://d3js.org/). In each phase, users were presented with instructions, from where they could start the test in normal mode, or sandbox mode. The latter allowed users to execute every test without registering any data. Finally, in each phase, users were permanently shown their keys and respective categories.

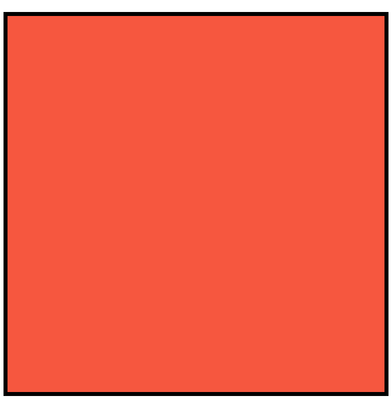


Figure 6: Example of an encoded value for the color luminance primitive visual task.

	Under 12	18-24	25-34	45-54	55-64
V. Position	2	12	4	4	1
H. Position	1	11	4	4	1
Length	2	11	5	3	2
Tilt	3	12	4	4	1
Size	2	12	4	3	2
Luminance	4	9	4	4	1

Table 1: Users' age distribution per primitive visual tasks.

3.4 Users

We gathered 24 users and asked them to execute all six tests (one per primitive visual task). However, not all tests were performed by all of them, mainly because they felt they were getting tired. We decided not to discard results because the information was still valid, and our statistic tests could still be applied. Besides, since each primary task was executed by all users, or by a subset of them, our study design remained within-subjects. Since we designed each test to be independent, its data was saved, regardless of the tests' completion. We ended up with 21 users for the vertical position test, 22 users for the color luminance test, 24 for the line tilt test, and 23 users for the: horizontal position, line length, and area size tests. Users' age was between 'Under 12' and '55-64', and their gender was: male or female. A summary can be found in tables 1 and 2.

3.5 Data Analysis

As mentioned before, we ended with six different datasets, one per test. First, we executed a statistical analysis on each one of them to assess how primitive visual tasks evolved between phases, and then we compared the best phases between each other. Our study design was a within-subjects. We studied one independent variable (the phase), which had three levels (1, 2, and 3). The dependent variable was response time (continuous) and accuracy (nominal or ordinal).

We studied accuracy in two ways, each aiming at different insights. In each phase of each test, there were always boundaries separating each category. Of the 12 values, a few were closer to the boundaries than the others. Closer values were more difficult to separate between categories.

In the nominal approach, we just considered answers to be right (1) or wrong (0), while in the ordinal approach, we considered answers to be right (1), or wrong ([0,1[). For the ordinal approach, we created a function to calculate accuracy, depending on the current phase. For example, in the first phase (two categories) of the horizontal position primitive visual tasks, six values were presented

	Female	Male	Other
V. Position	9	12	0
H. Position	9	14	0
Length	9	14	0
Tilt	9	15	0
Size	8	15	1
Luminance	9	13	0

Table 2: Users' gender distribution per primitive visual tasks.

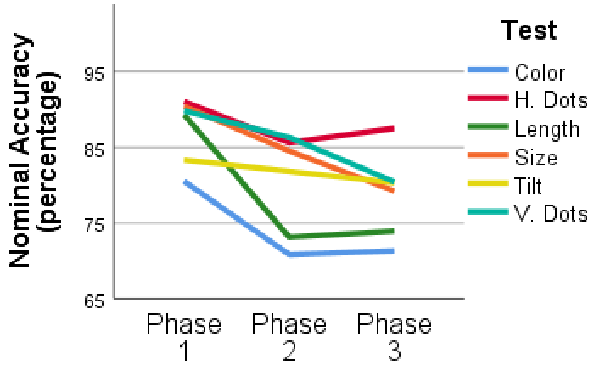


Figure 7: Nominal accuracy in the several primitive visual tasks.

Null Hypothesis			
The distributions of phases 1, 2, and 3 are the same.			
Test			
Cochran's Q Test			
Task	**Statistic**	**Significance**	**Decision**
Vertical	9.816	.007	Reject
Horizontal	3.108	.221	Retain
Length	22.917	<.0005	Reject
Tilt	.339	.844	Retain
Size	10.263	.006	Reject
Luminance	4.617	.099	Retain

Table 3: Nominal accuracy hypothesis tests summary.

in the left category, and six values in the right category. If a value appeared on the left side, but the user answered it was on the right side, the function calculated the distance between the value presented and the closest boundary of the chosen category. In this case, if the value presented was the nearest, the accuracy value would be 5/6 because there are six values on the left side. In summary, the possible accuracy values in the first phase were: 0, 1/6, 2/6, 3/6, 4/6, and 5/6. Then, in the second phase (three categories, each with four values in each), the possible accuracy values were: 0, 1/4, 2/4, and 3/4. Finally, in the third phase (four categories, each with 3 values in each) they were: 0, 1/3, and 2/3.

4 RESULTS

Response time was a continuous variable, but since normality could not be assumed according to the Shapiro-Wilk test, we used Friedman's non-parametric test. For nominal accuracy, we used the Cochran's Q test. For ordinal accuracy, we also used the Friedman's test.

4.1 Nominal Accuracy

In each test, Cochran's Q test [7] was run to assess if the percentage of users answering correctly was different between the phases. The sample size was adequate to use the X^2-distribution approximation [29]. Pairwise comparisons were performed using Dunn's [10] procedure with a Bonferroni correction for multiple comparisons. Adjusted p-values are presented:

Vertical Position. Accuracy was statistically significantly different between the three phases (Table 3); it decreased significantly between phases 1, and 3 (p = .008) (Figure 7). Therefore, choosing between two categories works best when finding vertical positions.

Horizontal Position. Accuracy was not statistically significantly different between the three phases (Table 3) (Figure 7). Thus, choosing between two, three, or four categories is similarly accurate when finding horizontal positions.

Line Length. Accuracy was statistically significantly different between the three phases (Table 3); it decreased significantly from phases 1 and 2 (p < .0005), and phases 1 and 3 (p < .0005) (Figure 7). So, choosing between two categories works best when finding length.

Line Tilt. Accuracy was not statistically significantly different between the three phases (Table 3) (Figure 7). Therefore choosing between two, three, or four categories is similarly accurate when finding tilt.

Area Size. Accuracy was statistically significantly different between the three phases (Table 3); it decreased significantly from phases 1 and 3 (p = .004) (Figure 7). Thus, choosing between two categories works best when finding size.

Color Luminance. Accuracy was not statistically significantly different between the three phases (Table 3) (Figure 7). So, choosing between two, three, or four categories is similarly accurate when finding color luminance.

Between Primitive Visual Tasks. In every test, the first phase was always the best one (Figure 7). Accuracy was statistically significantly different between every test, $X^2(5) = 27.904, p < .0005$. The Color Luminance test was the only one showing significant differences; with Horizontal Position (p < .0005), Vertical Position (p = .001), Line Length (p = .010), and Area Size (p = .004). Therefore with two categories, every task is similarly accurate, except finding color luminance.

4.2 Ordinal Accuracy

For each task, a Friedman test was run to assess if there were differences in accuracy between phases. Pairwise (Wilcoxon) comparisons were performed with a Bonferroni correction for multiple comparisons:

Vertical Position. Accuracy was statistically significantly different between the three phases (Table 4). However, Post hoc analysis showed no statistically significant differences between each pair of phases. Thus, choosing between two, three, or four categories is similarly accurate to find a dot's vertical position.

Horizontal Position. Accuracy was not statistically significantly different between the three phases (Table 4). So, choosing between two, three, or four categories is similarly accurate to find a dot's horizontal position.

Line Length. Accuracy was statistically significantly different between the three phases (Table 4). Post hoc analysis showed statistically significant differences between phases 1 and 2 (p = .037), 1 and 3 (p = .029). Therefore, choosing between two categories is more accurate to find a line's length.

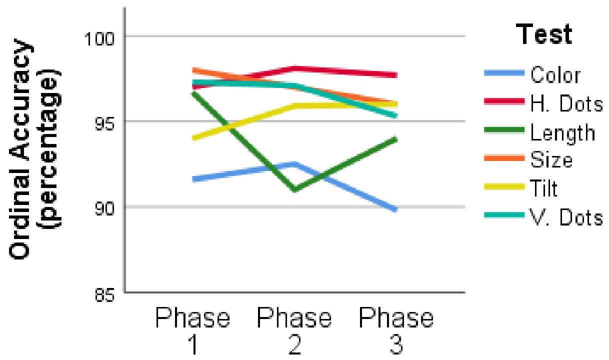


Figure 8: Ordinal accuracy in the several primitive visual tasks.

Line Tilt. Accuracy was not statistically significantly different between the three phases (Table 4). Thus, choosing between two, three, or four categories is similarly accurate to find a line's length.

Area Size. Accuracy was statistically significantly different between the three phases (Table 4). However, Post hoc analysis showed no statistically significant differences between each pair of phases. So, choosing between two, three, or four categories is similarly accurate to find an area's size.

Color Luminance. Accuracy was not statistically significantly different between the three phases (Table 4). Therefore, choosing between two, three, or four categories is similarly accurate to find a area's color luminance.

Between Tasks. In each task, the median in each phase was always 1.000, so, to compare them between each other, we choose each first phase. Accuracy was statistically significantly different between every tasks $X^2(2) = 30.588, p < .0005$. However, Post hoc analysis showed no statistically significant differences between each pair of tasks. Thus, choosing between tasks is similarly accurate (means for every test, per task, in figure 8).

4.3 Response Time

For each task, a Friedman's test was run to decide if there were differences in the accuracy between the phases. Pairwise (Wilcoxon) comparisons were performed with a Bonferroni correction for multiple comparisons.

Null Hypothesis			
The distributions of phases 1, 2, and 3 are the same.			
Test			
Friedman's Two-Way Analysis of Variance by Ranks			
Task	**Statistic**	**Significance**	**Decision**
Vertical	11.149	.004	Reject
Horizontal	3.863	.145	Retain
Length	20.759	< .0005	Reject
Tilt	.882	.644	Retain
Size	13.348	.001	Reject
Luminance	5.269	.072	Retain

Table 4: Ordinal accuracy hypothesis tests summary.

Null Hypothesis			
The distributions of phases 1, 2, and 3 are the same.			
Test			
Friedman's Two-Way Analysis of Variance by Ranks			
Task	**Statistic**	**Significance**	**Decision**
Vertical	421.537	< .0005	Reject
Horizontal	458.000	< .0005	Reject
Length	436.000	< .0005	Reject
Tilt	432.170	< .0005	Reject
Size	444.000	< .0005	Reject
Luminance	378.594	< .0005	Reject

Table 5: Response time tests' summary.

Between Tasks. The results across every test were similar; response time always increased significantly with every phase; p-values were always less than .0005 (Table 5). The lower response times always happened during each first phase (Figure 9) (Table 6); this means that **choosing between two categories induces less cognitive load, in all tasks**. When we compared between tasks, response time was statistically significantly different $X^2(5) = 957.592, p < .0005$ (Figures 9). Post hoc analysis showed statistically significant differences between each pair of tasks; p-values were always less than .0005, except between the line length and vertical position tests. This means that **finding a dot's horizontal position induces less cognitive load than the other tasks**.

5 DISCUSSION

Our goal was to assess how accurate visualizations can be apprehended pre-attentively, and how much cognitive load they entail. We tested six primitive visual tasks, each in three phases (two, three, and four categories). Our method used the guided search theory, which states that each visual feature has several layers (categories). Accuracy was higher when users chose between two categories, which means they can pre-attentively see values if they only have to guess between two alternatives.

To measure accuracy, we used explicit testing with direct observation because users had to estimate each value's category, and with the method of constant stimuli from psychophysics to avoid biasing users with their given orders. Besides, unlike most studies, values were shown isolated, with no distractors because our goal is

Task	**Phase 1**	**Phase 2**	**Phase 3**
Vertical	.574s	.620s	.741s
Horizontal	.505s	.530s	.601s
Length	.593s	.677s	.783s
Tilt	.536s	.572s	.665
Size	.598s	.645s	.778s
Luminance	.651s	.717s	.759s

Table 6: Mean response times for every task, per phase.

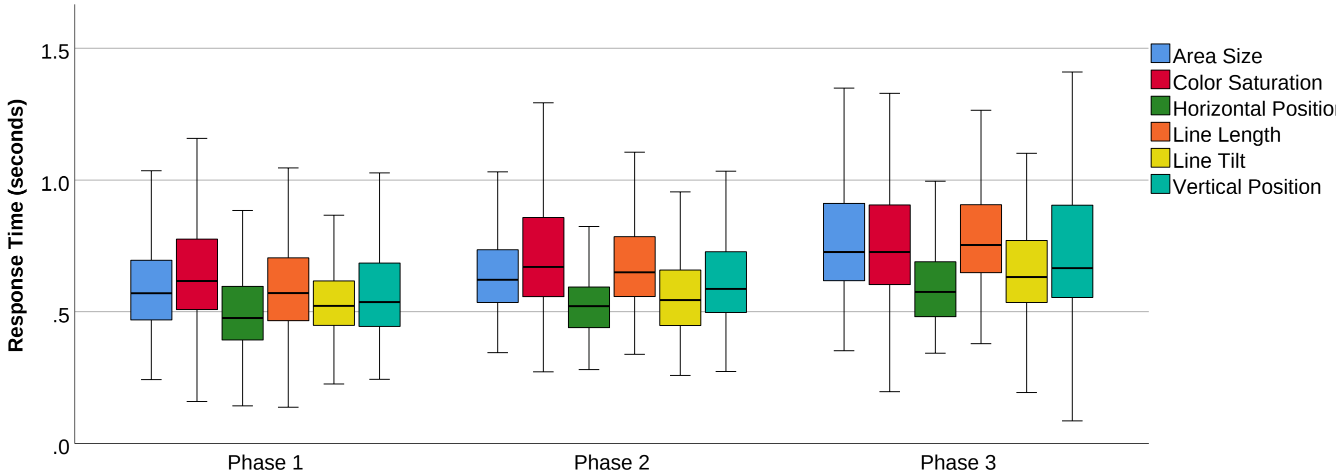


Figure 9: Response time in the several primitive visual tasks.

not making visual features pop-out, but to make each perception pre-attentive.

With a nominal accuracy (right or wrong) analysis, we concluded that finding horizontal positions is the most accurate primitive visual task and that finding color luminance intensities is the least accurate. These results are in line with earlier psychophysics studies' ranks (accuracy and difficulty) [4, 5, 19], of different perceptual evaluations. In them, the position task was on the first and second ranks, and color luminance in the sixth rank. However, these ranks are valid for comparisons and quantitative evaluations during our attentive phase of visualizations with over one encoded value at a time.

With an ordinal accuracy analysis, results showed that we can pre-attentively execute each of our six tested primitive visual tasks. We got no significant differences between tasks, which means they are as accurate. This happened because we measured accuracy according to each answer's difficulty. The majority of wrong answers were given when values got encoded near borders. Although this does not change the percentage of wrong answers, it still shows that these primitive visual tasks can still be executed pre-attentively. So, one key design solution to present values is to have in account the number of ranges chosen. If users are requested to guess between four categories, then the visualization should only present four values, each in the middle of its corresponding bin. For example, with four categories, where the values vary between 0 and 400, the set of values would be 50, 150, 250, and 350.

In the end, we measured the cognitive load by analyzing the mean response time for each task. Results were as with the nominal accuracy: the best phase in every task was the first one, the fastest primitive visual task was finding horizontal positions, and the slowest was finding color luminance intensities. However, this time our results were significantly different between each pair of tasks, which means we could create a rank for every task. So, combining both accuracy and response time, we may conclude that finding horizontal positions is indeed the most accurate pre-attentive primitive visual task.

6 DESIGN IMPLICATIONS

If designers want visualizations to be perceived pre-attentively, they should choose specific primitive visual tasks. Overall, the visualizations we tested proved to be accurate without demanding too much cognitive load. Here are the most important guidelines:

- visualizations should encode values using horizontal position variations, if designers want to achieve faster response times, and higher perception accuracy;
- the number of values a visualization should encode, depends on the number of bins created in it, and users can accurately perceive values up to four bins.
- differentiating between different luminances and lengths is not advised in pre-attentive visualizations.

7 CONCLUSION

Incidental visualizations are within our grasp; our study allowed us to understand which primitive visual tasks are more accurate and faster to execute. Finding horizontal positions in visualizations proved to be the best primitive visual task, and color luminance identification the worst one. We hope that our guidelines can aid researchers and developers to create and further investigate new visualizations. For future work, we think it is important to explore each primitive visual task individually to assess the limits of users' perception. In particular, how many bins (categories) can users differentiate, until their perceptions stop being pre-attentive? Maybe the key to incidental visualizations is to understand the best discretization for each primitive primary task.

ACKNOWLEDGMENTS

This work was supported by FCT, through grants VisBig PTDC/CCI-CIF/28939/2017, UIDB/50021/2020, and SFRH/BD/143496/2019.

REFERENCES

[1] T. S. Amer. 2005. Bias Due to Visual Illusion in the Graphical Presentation of Accounting Information. *Journal of Information Systems* 19, 1 (mar 2005), 1–18. https://doi.org/10.2308/jis.2005.19.1.1

[2] Tara C. Callaghan. 1984. Dimensional interaction of hue and brightness in preattentive field segregation. *Perception & Psychophysics* 36, 1 (jan 1984), 25–34. https://doi.org/10.3758/bf03206351

[3] C. Melody Carswell and Christopher D. Wickens. 1987. Information integration and the object display An interaction of task demands and display superiority. *Ergonomics* 30, 3 (mar 1987), 511–527. https://doi.org/10.1080/00140138708969741

[4] William S. Cleveland and Robert McGill. 1984. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *J. Amer. Statist. Assoc.* 79, 387 (sep 1984), 531–554. https://doi.org/10.1080/01621459.1984.10478080

[5] William S. Cleveland and Robert McGill. 1985. Graphical Perception and Graphical Methods for Analyzing Scientific Data. *Science* 229, 4716 (1985), 828–833. http://www.jstor.org/stable/1695272

[6] William S. Cleveland and Robert McGill. 1987. Graphical Perception: The Visual Decoding of Quantitative Information on Graphical Displays of Data. *Journal of the Royal Statistical Society. Series A (General)* 150, 3 (1987), 192. https://doi.org/10.2307/2981473

[7] W. G. Cochran. 1950. The Comparison of Percentages in Matched Samples. *Biometrika* 37, 3/4 (dec 1950), 256. https://doi.org/10.2307/2332378

[8] Frederick E. Croxton and Harold Stein. 1932. Graphic Comparisons by Bars, Squares, Circles, and Cubes. *J. Amer. Statist. Assoc.* 27, 177 (mar 1932), 54–60. https://doi.org/10.1080/01621459.1932.10503227

[9] Frederick E. Croxton and Roy E. Stryker. 1927. Bar Charts versus Circle Diagrams. *J. Amer. Statist. Assoc.* 22, 160 (dec 1927), 473–482. https://doi.org/10.1080/01621459.1927.10502976

[10] Olive Jean Dunn. 1964. Multiple Comparisons Using Rank Sums. *Technometrics* 6, 3 (aug 1964), 241–252. https://doi.org/10.1080/00401706.1964.10490181

[11] Richard Dunn. 1988. Framed Rectangle Charts or Statistical Maps with Shading: An Experiment in Graphical Perception. *The American Statistician* 42, 2 (may 1988), 123–129. https://doi.org/10.1080/00031305.1988.10475541

[12] Walter Crosby Eells. 1926. The Relative Merits of Circles and Bars for Representing Component Parts. *J. Amer. Statist. Assoc.* 21, 154 (jun 1926), 119–132. https://doi.org/10.1080/01621459.1926.10502165

[13] Sara Irina Fabrikant, Stacy Rebich Hespanha, and Mary Hegarty. 2010. Cognitively Inspired and Perceptually Salient Graphic Displays for Efficient Spatial Inference Making. *Annals of the Association of American Geographers* 100, 1 (jan 2010), 13–29. https://doi.org/10.1080/00045600903362378

[14] Joseph H. Goldberg and Jonathan I. Helfman. 2010. Comparing information graphics. In *Proceedings of the 3rd BELIV'10 Workshop on BEyond time and errors: novel evaLuation methods for Information Visualization - BELIV '10*. ACM Press. https://doi.org/10.1145/2110192.2110203

[15] C.G. Healey and J.T. Enns. 1999. Large datasets at a glance: combining textures and colors in scientific visualization. *IEEE Transactions on Visualization and Computer Graphics* 5, 2 (1999), 145–167. https://doi.org/10.1109/2945.773807

[16] Christopher G. Healey, Kellogg S. Booth, and James T. Enns. 1996. High-speed visual estimation using preattentive processing. *ACM Transactions on Computer-Human Interaction* 3, 2 (jun 1996), 107–135. https://doi.org/10.1145/230562.230563

[17] Brian M. Hughes. 2001. Just Noticeable Differences in 2D and 3D Bar Charts: A Psychophysical Analysis of Chart Readability. *Perceptual and Motor Skills* 92, 2 (apr 2001), 495–503. https://doi.org/10.2466/pms.2001.92.2.495

[18] Gordon E. Legge, Yuanchao Gu, and Andrew Luebker. 1989. Efficiency of graphical perception. *Perception & Psychophysics* 46, 4 (jul 1989), 365–374. https://doi.org/10.3758/bf03204990

[19] STEPHAN LEWANDOWSKY and IAN SPENCE. 1989. The Perception of Statistical Graphs. *Sociological Methods & Research* 18, 2-3 (nov 1989), 200–242. https://doi.org/10.1177/0049124189018002002

[20] Rudolf Netzel, Jenny Vuong, Ulrich Engelke, Seán O'Donoghue, Daniel Weiskopf, and Julian Heinrich. 2017. Comparative eye-tracking evaluation of scatterplots and parallel coordinates. *Visual Informatics* 1, 2 (jun 2017), 118–131. https://doi.org/10.1016/j.visinf.2017.11.001

[21] C. North. 2006. Toward measuring visualization insight. *IEEE Computer Graphics and Applications* 26, 3 (may 2006), 6–9. https://doi.org/10.1109/mcg.2006.70

[22] Lewis V. Peterson and Wilbur Schramm. 1954. How accurately are different kinds of graphs read? *Educational Technology Research and Development* 2, 3 (jun 1954), 178–189. https://doi.org/10.1007/bf02713334

[23] Young Sam Ryu, Beth Yost, Gregorio Convertino, Jian Chen, and Chris North. 2003. Exploring Cognitive Strategies for Integrating Multiple-View Visualizations. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 47, 3 (oct 2003), 591–595. https://doi.org/10.1177/154193120304700371

[24] Priti Shah and Patricia A. Carpenter. 1995. Conceptual limitations in comprehending line graphs. *Journal of Experimental Psychology: General* 124, 1 (1995), 43–61. https://doi.org/10.1037/0096-3445.124.1.43

[25] Daniel J. Simons and Daniel T. Levin. 1997. Change blindness. *Trends in Cognitive Sciences* 1, 7 (oct 1997), 261–267. https://doi.org/10.1016/s1364-6613(97)01080-2

[26] Daniel J. Simons and Ronald A. Rensink. 2005. Change blindness: past, present, and future. *Trends in Cognitive Sciences* 9, 1 (jan 2005), 16–20. https://doi.org/10.1016/j.tics.2004.11.006

[27] Ian Spence. 1990. Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance* 16, 4 (1990), 683–692. https://doi.org/10.1037/0096-1523.16.4.683

[28] Joseph K. Tan. 1994. Human processing of two-dimensional graphics: Information-volume concepts and effects in graph-task fit anchoring frameworks. *International Journal of Human-Computer Interaction* 6, 4 (oct 1994), 414–456. https://doi.org/10.1080/10447319409526104

[29] Merle W. Tate and Sara M. Brown. 1970. Note on the Cochran Q Test. *J. Amer. Statist. Assoc.* 65, 329 (mar 1970), 155–160. https://doi.org/10.1080/01621459.1970.10481069

[30] Martha Teghtsoonian. 1965. The Judgment of Size. *The American Journal of Psychology* 78, 3 (sep 1965), 392. https://doi.org/10.2307/1420573

[31] Anne M. Treisman and Garry Gelade. 1980. A feature-integration theory of attention. *Cognitive Psychology* 12, 1 (jan 1980), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

[32] Edward R. Tufte. 1992. *The Visual Display of Quantitative Information*. Graphics Pr. https://www.amazon.com/Visual-Display-Quantitative-Information/dp/096139210X?SubscriptionId=AKIAIOBINVZYXZQZ2U3A&tag=chimbori05-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=096139210X

[33] Susan Vanderplas, Dianne Cook, and Heike Hofmann. 2020. Testing Statistical Charts: What Makes a Good Graph? *Annual Review of Statistics and Its Application* 7, 1 (jan 2020). https://doi.org/10.1146/annurev-statistics-031219-041252

[34] Susan VanderPlas and Heike Hofmann. 2015. Signs of the Sine Illusion—Why We Need to Care. *Journal of Computational and Graphical Statistics* 24, 4 (oct 2015), 1170–1190. https://doi.org/10.1080/10618600.2014.951547

[35] Susan VanderPlas and Heike Hofmann. 2016. Clusters Beat Trend!? Testing Feature Hierarchy in Statistical Graphics. *Journal of Computational and Graphical Statistics* 26, 2 (jul 2016), 231–242. https://doi.org/10.1080/10618600.2016.1209116

[36] Susan VanderPlas, Goluch C Ryan, and Heike Hofmann. 2019. Framed! Reproducing and Revisiting 150-Year-Old Charts. *Journal of Computational and Graphical Statistics* 28, 3 (apr 2019), 620–634. https://doi.org/10.1080/10618600.2018.1562937

[37] Johan Wagemans, James H. Elder, Michael Kubovy, Stephen E. Palmer, Mary A. Peterson, Manish Singh, and Rüdiger von der Heydt. 2012. A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin* 138, 6 (2012), 1172–1217. https://doi.org/10.1037/a0029333

[38] Johan Wagemans, Jacob Feldman, Sergei Gepshtein, Ruth Kimchi, James R. Pomerantz, Peter A. van der Helm, and Cees van Leeuwen. 2012. A century of Gestalt psychology in visual perception: II. Conceptual and theoretical foundations. *Psychological Bulletin* 138, 6 (2012), 1218–1252. https://doi.org/10.1037/a0029334

[39] Jeremy M. Wolfe. 1994. Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review* 1, 2 (jun 1994), 202–238. https://doi.org/10.3758/bf03200774

[40] Emanuel Zgraggen, Zheguang Zhao, Robert Zeleznik, and Tim Kraska. 2018. Investigating the Effect of the Multiple Comparisons Problem in Visual Analysis. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press. https://doi.org/10.1145/3173574.3174053