Exploring how Temporal Framing Affects Trust with Time-series Visualizations

Tomás Alves, Carlota Dias, Daniel Gonçalves, Sandra Gama
HUMAN Lab, Graphics and Interaction, INESC-ID, IST,
University of Lisbon, Portugal
{tomas.alves,carlota.lopes.dias,daniel.j.goncalves,sandra.gama}@tecnico.ulisboa.pt

***Abstract*—Trust is one of the most relevant factors when users build knowledge from visualization to predict whether they will use the represented information. In particular, trust perception is the user's subjective evaluation of the quality and reliability of the visualized information. However, research leveraging information visualization techniques to study trust perception is limited. This work studies whether varying the temporal framing of line charts affects trust perception in an uncertain scenario. Our results suggest that granularity may be relevant for time-based visualization design. In particular, individuals trust more in a line chart with a higher number of data points and interact more with a line chart in which they trust less. These findings contribute to the state-of-the-art research in visual analytic systems by empowering designers to understand how trust perception in a health emergency scenario varies for line charts with different temporal frames.**

***Index Terms*—information visualization, trust, user interaction, time-series**

I. Introduction

The core purpose of an Information Visualization (InfoVis) is to help users discover, explain, and form decisions based on the information that is conveyed [1]. Visualizations may follow several strategies to aid decision-making, e.g., offering sufficient guidance [2] or emphasizing critical information [3]. However, InfoVis research shows that visualization-supported decision-making vulnerable to cognitive biases and uncertainty [4], [5] Additionally, other human factors affect how the user derives a choice of direction as an outcome. For instance, when users interact with a visualization, the level of trust that users have in a new knowledge affects their decision-making process [6].

In InfoVis, trust defines the tendency of a user to rely on visualization and build on the information displayed [7]. In particular, trust becomes particularly important when there is some risk associated with the information; it allows the user to minimize the uncertainty belonging to digital data, especially when they are vulnerable to suffering a loss if they believe in the information displayed [7]. However, trust remains a challenge in InfoVis design [8] since there is still limited evidence regarding what might lead a user to trust in visualization without an extensive elaboration of the information [7].

In our study, we focus on trust perception in an uncertainty scenario. Further, we can assume that when users are prompt to decide with risk, they will elaborate on the trustworthiness of the visualization more in-depth yet rely on superficial trust cues in less relevant or less risky tasks [7]. Inspired by these findings, we place subjects in a decision-making situation with an involved risk. At the same time, this situation is familiar and relatable enough to emphasize the risky nature of the decision. In particular, we use a health emergency scenario when there is an overcrowding crisis, i.e., there is no space left to meet the timely needs of the following patient requiring emergency care [9]. We opted for this topic because we believe that most people are familiar with having to wait for medical support in the emergency department. Then, we ask users to make decisions and perform tasks based on time-oriented information visualizations with varying domain framing factors. The study aimed to address questions such as: Does the temporal frame of the time series affect trust perception? Does the degree of confidence in a choice vary across several decisions? Do users interact more with a visualization they trust the most?

II. Related Work

The topic of trust has been relatively underexplored in visual analytics. Similar to human relations where there are both a trustor and a trustee, trust in the InfoVis context encompasses trustworthiness and trust perception [7]. The trustworthiness of the visualization depends on the characteristics of the visualization like data accuracy, objectivity, and completeness [10]. For instance, Xiong et al. [10] explored whether there is a relationship between trust and data visualization transparency – the perceived quality and quantity of intentionally shared information [11]. Xiong et al. [10] asked participants to put themselves in the role of a firefighter to promote a frame of reference. Then, each participant chose which visualization to use and from which fire station they should dispatch the firefighters. Participants were shown two different visualizations and then told these visualizations were screenshots of several driving applications that displayed the routes from several fire stations to a fire location. These visualizations varied in the displayed volume of information by changing the number of possible routes, the number of fire stations recommended, and the number of fastest paths. Results showed that participants were more likely to choose visualizations that appeared to be clear, more thorough, and disclosed a higher amount of information.

Other studies have shown that design factors such as usability and user experience or the amount of processed underlying data can affect trustworthiness [12], [13]. For instance, a recent study by Bartram et al. [14] shows that trust plays a significant role in data workers. In particular, these workers would be willing to perform monotonous and repetitive tasks to maintain immediate access, control, and understanding over their actions and sense-making process.

Regarding trust perception, it tackles the evaluation of the quality and reliability of the visualized information [7]. While Kong et al. [13] suggest that the misalignment of graphical elements affects the understanding and, consequently, the credibility of the information depicted, other researchers imply that prior experiences play a relevant role [15], [16]. For instance, Dasgupta et al. [17] studied the level of trust of domain scientists in visual analytics systems as opposed to more common manual analysis methods. The authors were able to find that, despite being unfamiliar with a visual analytic system, the experts had an average level of trust comparable with the same in conventional analysis methods. The core factors for the analytical system are that it should be intuitive, transparent, and allow a seamless switch between hypothesis generation and evidence gathering. Finally, user intentions and perceived risk may also influence trust perception [18], [19]. These studies collectively show that studying trust in visualization offers an opportunity for the state-of-the-art. Our work builds on the mentioned studies for trust assessment, focusing on whether temporal dimensions affect trust perception.

III. Methodology

Our research question focuses on analyzing **whether temporal framing has an effect on trust perception in the context of time-oriented linear charts of healthcare information**. We leverage the self-assessment of trust and their interaction data with the information visualizations.

A. Temporal Framing

There is a wide range of data features to consider when studying trust perception. Time dimensionality [20] may hold promising results given its relevance in recent research [21], [22]. Time-oriented data includes dimensions such as linear $vs.$ cyclic time, time points $vs.$ time intervals, and order time $vs.$ branching time. Consequently, different time features lead to alternative visualization techniques. Among the different time-based aspects of interest, we believe that *temporal framing*, i.e., the temporal scope that is presented in the domain of visualization, such as daily, weekly, or monthly, may hold promising results. This feature can offer several factors to manipulate, e.g., time range or value aggregation. Therefore, temporal framing may significantly affect the level of detail of the information and, consequently, the granularity of the data or the number of data points.

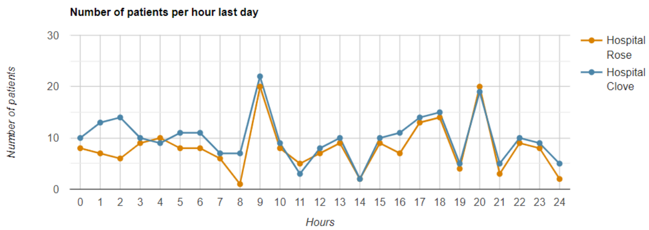
For instance, Oscar et al. [23] studied the consequences of mismatching the granularity of information presented on visualization to user needs. Participants were shown different visualizations and asked to complete tasks that required information that might not be available in the visualization granularity. In particular, in some cases, users were not presented with enough information to answer the questions. Results showed that when users analyzed information mismatched to the need for detail required by the task, they were less likely to complete the assignment correctly. Moreover, participants were often unable to identify that the visualization did not include the information required to complete the task. Consequently, results demonstrated that using an appropriate visualization is a crucial performance factor. Although the mentioned studies recognize that different visual techniques affect the decisions users make when analyzing distinct visualizations, as far as we know, there is no information concerning how using time-series visualization techniques impacts users' perceived trust.

B. Visualizations

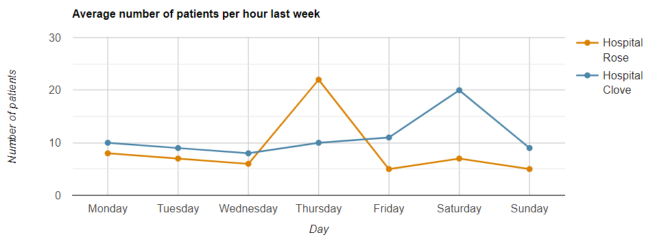
Research has shown that positional encodings such as line charts are the best option to visualize time-oriented data on decision-making processes [21], [24]. Moreover, line charts are one of the most well-known visualization idioms, and familiarity with a visualization system inspires trust, whereas novel visualizations may act as a barrier [15]. Based on these findings, we opted to use line charts to depict the variation of a continuous variable along a time axis (Figure 1), as this type of graph is the most common form of time-series visualization [25]. Each line chart displays the number of patients in an emergency room throughout a time in two different hospitals, encoded by two colors randomly associated at the beginning of each experiment. While the x-axis illustrates the time, the y-axis represents the number of patients, and each line presents a hospital. Each line chart has one of the following possible temporal framing values: (i) the number of patients per hour in the past day (V-Day), (ii) the average number of patients per hour in the past week (V-Week), and (iii) the average number of patients per hour in the past month (V-Month). Moreover, Kong et al. [13] point out that users elaborate on information more or less deeply to decide whether it is trustworthy based on different situative factors. To avoid introducing some bias, we create each visualization from the same database to keep data trustworthiness equal across all conditions, i.e., the weekly data is a subset of the monthly data since the former reports the last seven days of the latter. Therefore, the visualizations contain all information the user needs to evaluate the quality of the underlying data.

C. Tasks

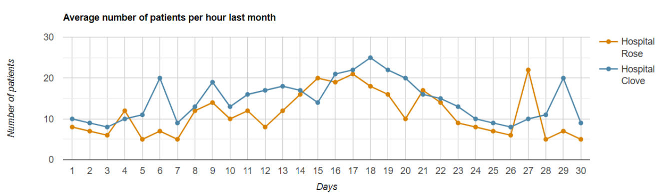
Our study includes a sequence of decision tasks (Figure 2). Firstly, we focus on the self-calibrated degree of confidence in a taken decision [17] by asking participants to put themselves in a position of having a health emergency and needing to decide to which hospital they should go to (Figure 2, left). Then, we asked each participant to assess the three visualizations simultaneously and to choose between the two hospitals which one they would like to go to using a think-aloud protocol. We prompted subjects to decide solely based on the number of patients that visited the emergency department



(a) The number of patients per hour in the past day.



(b) The average number of patients per hour in the past week.



(c) The average number of patients per hour in the past month.

Fig. 1. Set of visualizations with different time granularity factors.

on the past day, week, or month. We asked the testers to consider that fewer patients would most likely lead to a shorter waiting time. We made up the names of the hospitals, and we randomly assigned the color encoding of each hospital to reduce any potential biases. As we mentioned, we kept the same number of patients in each of the two hospitals between the visualizations. Therefore, we prompt the subject to base their decision on the temporal framing factor. Moreover, we asked each participant to report which of the line charts is their *anchor frame*, i.e., which frame weighted the most on their decision to choose between the two data trends.

Secondly, participants were assigned a random order through which they would interact with each visualization separately (Figure 2, middle). We asked participants to complete three tasks to ensure that they acknowledged the different framing for each visualization. The tasks consisted of (i) finding the number of patients for a specific point in time, (ii) finding what hospital had the greatest growth of patients in a specific time interval, and (iii) choosing one hospital to visit in case of an emergency. After performing the tasks, we invited users to assess the trustworthiness of the visualization with that they interacted.

For the last part, we presented again to each subject the three visualizations simultaneously (Figure 2, right). We then asked participants to choose which hospital they would go to and which visualization weighted the most in their decision, similar to the first part. This repetition of the first part may allow us to understand the self-calibrated degree of confidence in a taken decision [17], i.e., whether performing tasks with each visualization may alter the choice that the participant made at the beginning. By making participants explore the visualizations in more detail, we believe their perceived trustworthiness towards the same visualizations may change. Therefore, the last part of our experience allows us to understand whether an extensive individual analysis of each visualization affects the interaction and decision-making patterns when they must again report the frame they rely on the most to choose between hospitals. In particular, this three-step methodology supports the analysis of whether there is any specific temporal framing that participants trust the most and if that temporal framing functions as an anchor to the overall decision-making.

D. Measures

Demographics We recorded the gender, age, self-reported visual acuity of each participant, and whether they were color-blind by a validated simplified version of the Ishihara test [26]. We also controlled other external factors such as the last time the participant visited a hospital for an emergency (last week, last month, last year, or never), and their familiarity with line charts using a five-point Likert scale ranging from *"Not familiarized"* to *"Completely familiarized"*.

Trust Assessment We assessed visualization trustworthiness with a five-point Likert scale ranging from *"I do not trust this information"* (1) to *"I completely trust this information"* (5) for each visualization framing.

User interaction We measure the **response time** for each tested task in seconds. In addition, we assess the **number of hover events per data point** that the participant triggered while interacting with each graph. A hover event is triggered when the user hovers over a data point to inspect the number of patients through a tooltip.

E. Procedure

We recruited participants through standard convenience sampling procedures by direct contact and word of mouth. Our final data set is composed of 89 participants (38 males, 51 females) between 18 and 69 years old ($M = 27.40; SD = 12.04$). Participants are general end-users with no particular relation to the healthcare area and with normal or corrected-to-normal vision. User tests were conducted through an online videoconference platform, forcing the visualizations to resize to ensure it was displayed in the same physical size, regardless of device resolution. After participants provided informed consent, we first asked them to read a document that introduces the context for the visualizations and prompts participants to be aware of the emergency associated with this crisis while motivating trustworthiness. In addition, the document explains the context of the data of the visualizations and the negative consequences of overcrowding in emergency departments, e.g., the increase in mortality rates.

We continued in a three-part test, as depicted in Figure 2. First, we asked participants to put themselves in a position

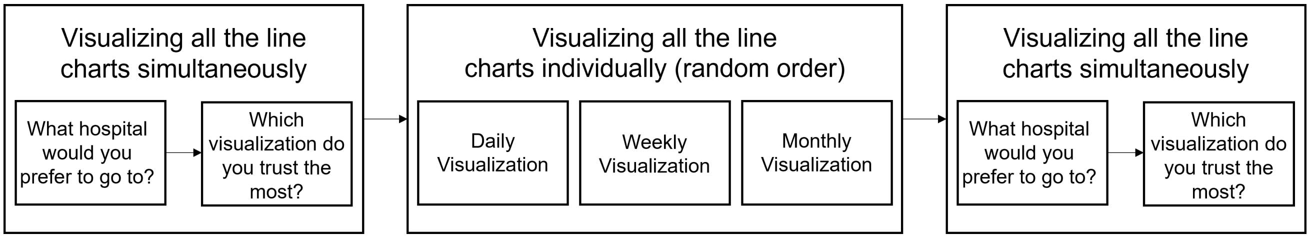


Fig. 2. Overview of the decision processes in our experiment.

of having a health emergency and having the need to decide which hospital they should go to. Participants decided while seeing the three visualizations simultaneously. Then, we asked them to choose which hospital they would prefer to go to and which visualization they trusted the most to make that decision. The assistant collected the time participants took to decide and registered their anchor frame. Next, participants performed the three mentioned tasks separately in random order in each visualization. Additionally, we invited subjects to assess their perceived trust regarding each visualization after interacting with it. The assistant collected the number of hovers that subjects triggered while interacting with each line chart and the time to complete each task. Afterward, we presented to each subject the three visualizations simultaneously. Then, we asked them to choose which hospital they would go to and their anchor frame. Finally, each subject filled in the demographic questionnaire.

F. Research Design and Data Analysis

We ran one-way ANOVAs with the temporal frame (3 levels) as the independent factor to study its effect on the dependent variables (trust perception, decision time, and hovers per point). We also ran two-way mixed ANOVAs with anchor frame (3 levels) and temporal frame (3 levels) as independent factors to understand whether the anchor frame plays a role in these relationships. As we mentioned, the anchor frame is a between-subjects variable and has three possible values: {A-Day, A-Week, A-Month}. The temporal frame is a within-subjects variable with three possible values as well: {V-Day, V-Week, V-Month}. All evaluation sessions were video-recorded to collect interaction metrics. We measured user interaction through two variables: *number of hovers per point*, represented by the sum of hover events that the participant triggered in a visualization divided by the number of data points; and *time to choose a hospital*, which corresponds to the time users take to pick a hospital while analyzing the visualizations individually. We tested for sphericity (Mauchly's test) and used the Greenhouse-Geisser correction when the assumption was not met. ANOVAs were followed by posthoc Tukey's range tests, which include Bonferroni corrections. Finally, we examine whether participants changed their anchor frame between the choice moments. We ran a chi-square test of independence for $r \times c$ contingency tables.

IV. Results

The following subsections discuss the results regarding the self-assessment of trust perception and user interaction metrics. Data are mean $\pm$ standard error unless otherwise stated.

A. Trust Perception

We started studying whether trust perception was influenced by the temporal framing when participants analyzed the different visualizations separately. Results did not show a statistically significant interaction between the framing and perceived trust, $F(1.636, 143.971) = 2.777$, $p = .076$, partial $\eta^2 = .031$. All distributions look similar with a positive mean rating, yet results suggest that subjects assessed their trust perception with lower grades in V-Week $(4.056 \pm .091)$ compared to V-Day $(4.135 \pm .096)$ and V-Month $(4.213 \pm .088)$ granularity values (Figure 3). In particular, a pairwise comparison reports a statistically significant increase of 0.157 (95% CI, 0.002 to 0.313) points from V-Week to V-Month, $p = .046$.

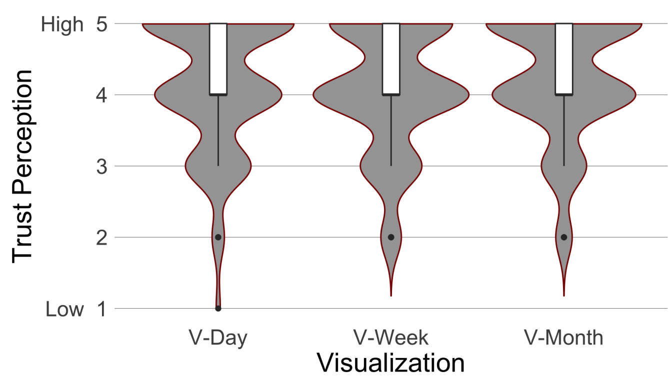


Fig. 3. Violin and boxplots of trust perception for each visualization.

Next, we ran a two-way mixed ANOVA to study whether the anchor frame affects the degree of trust perception per temporal framing (Figure 4). There was a statistically significant interaction between the anchor frame and the temporal framing on trust perception, $F(3.372, 144.982) = 4.469$, $p = .003$, partial $\eta^2 = .094$. In a pairwise comparison, we can observe that, for subjects that chose A-Month, there were statistically significant increases of 0.275 (95% CI, 0.027 to 0.522) points from V-Day to V-Month, $p = .024$, and of 0.333 (95% CI, 0.140 to 0.527) points from V-Week to V-Month, $p < .001$.

These results suggest that people who rely more on A-Month consistently assess their perceived trust in the remaining granularity options (V-Day: $4.020 \pm .127$; V-Week: $3.961 \pm .120$; V-Month: $4.294 \pm .117$). Similarly, people who chose A-Day also trusted more V-Day ($4.417 \pm .261$) compared to V-Week ($4.250 \pm .248$) and V-Month ($4.000 \pm .241$) granularity values. In contrast, people who favored A-Week had similar values across their trust perception for V-Day ($4.231 \pm .178$), V-Week ($4.154 \pm .169$), and V-Month ($4.154 \pm .164$).

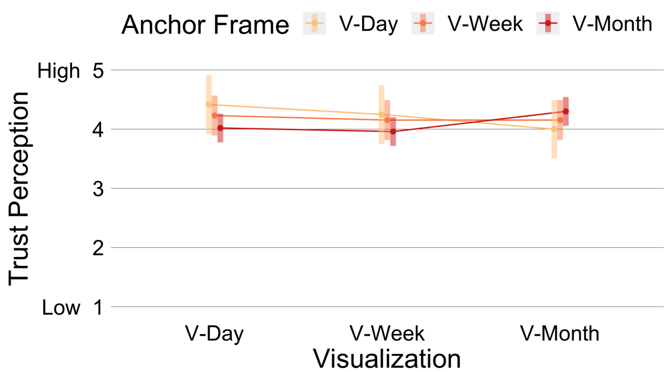


Fig. 4. Estimated marginal means of trust perception based on anchor granularity and the temporal framing of the assessed visualization.

We ran a chi-square test of independence between the anchor frame in both the initial and the final decision moments Table I). There was a statistically significant association between the chosen visualizations, $\chi^2(4) = 60.348, p < .001$. The association was moderately strong [27], Cramer's V $= 0.582$. Overall, these results suggest that participants were consistent in their choices, reflecting a robust self-calibrated degree of confidence in a decision, except for people who chose A-Week since they were more likely to change their decision after a careful analysis of the data. Moreover, A-Month was the most relied-on frame in both decision moments. Our next step is to find whether the interaction data reflects the lower trust perception of the V-Week line chart and how the anchor frame affected how people rated each visualization.

TABLE I
CROSSTABULATION OF THE ANCHOR FRAME IN EACH DECISION MOMENT. ADJUSTED RESIDUALS APPEAR IN PARENTHESES NEXT TO THE OBSERVED FREQUENCIES.

		Final Anchor Frame			
		A-Day	*A-Week*	*A-Month*	Total
Initial Anchor Frame	*A-Day*	8 (5.5)	0 (−1.9)	4 (−2.4)	12
	A-Week	3 (−0.5)	15 (5.4)	8 (−4.2)	26
	A-Month	2 (−3.3)	4 (−3.6)	45 (5.5)	51
	Total	13	19	57	89

B. User Interaction

We collected the time in seconds users took to decide between the two hospitals and how many times they hovered a point in the line charts. The temporal framing showed statistically significant effects both in the time users took to decide, $F(2, 174) = 8.221$, $p < .001$, partial $\eta^2 = .086$, as well as in how many hovers per point subjects did, $F(1.457, 126.670) = 23.251$, $p < .001$, partial $\eta^2 = .209$. Regarding the time to decide, participants took less time to choose a hospital when analyzing V-Day ($6.852 \pm .558$), compared to V-Month ($9.477 \pm .935$) and V-Week (11.068 ± 1.120) versions. More precisely, a pairwise comparison reported statistically significant increases of 4.216 (95% CI, 1.384 to 7.048) seconds from V-Day to V-Week, $p = .001$, and of 2.625 (95% CI, 0.354 to 4.891) seconds from V-Day to V-Month, $p = .017$. These results suggest that participants find it easier to decide when observing V-Day.

Results showed that participants performed more hover events per points when they analysed V-Week ($0.730 \pm .095$), followed by V-Day ($.307 \pm .038$) and then V-Month ($0.297 \pm .042$). In particular, a pairwise comparison reports statistically significant increases of 0.424 (95% CI, 0.219 to 0.629) hovers per point from V-Day to V-Week, $p < 0.001$, and of 0.434 (95% CI, 0.233 to 0.635) from V-Month to V-Week, $p < 0.001$. Therefore, users interacted more with V-Week, which was the one they rated with a lower trust perception.

Akin to the trust perception analysis, we decided to verify whether the anchor frame affects the time to decide on a hospital and the hovers per temporal framing. We found that the anchor frame significantly affected the time that people took to choose a hospital when they interacted with the visualizations one at a time (Figure 5), $F(3.706, 157.523) = 2.730$, $p = .035$, partial $\eta^2 = .060$. In particular, a pairwise comparison showed that subjects with A-Week had statistically significant increases of 5.385 (95% CI, 1.256 to 9.514) seconds from V-Day to V-Month, $p = .006$. Additionally, subjects that with A-Month showed statistically significant increases of 5.300 (95% CI, 1.540 to 9.060) seconds from V-Day to V-Week, $p = .003$, and of 3.900 (95% CI, 0.676 to 7.124) seconds from V-Month to V-Week, $p = .012$. These results show that people with either A-Day (11.000 ± 3.007) or A-Month (12.680 ± 1.473) take more time to choose a hospital when analyzing V-Week. Nevertheless, only the people who chose A-Day actually take less time to pick a hospital when they analyze their anchor visualization (5.917 ± 1.519). This trend is not present for people with other anchor frame values; those focused on A-Week (6.269 ± 1.032) or A-Month ($7.380 \pm .744$) are faster in visualization with V-Day.

Finally, we found that the anchor frame did not significantly affect the number of hovers per point performed by the participants when interacting with the visualizations individually (Figure 6), $F(2.864, 123.163) = 0.288$, $p = .825$, partial $\eta^2 = .007$. Even though there was non-significant interaction, a pairwise comparison showed that subjects with an A-Week had a statistically significant increase of 0.437 (95% CI, 0.053 to 0.820) hovers per point from V-Day to V-Week, $p = .020$. Moreover, subjects with A-Month showed statistically significant increases of 0.433 (95% CI, 0.159 to 0.707) hovers per points from V-Day to V-Week, $p = .001$, and of 0.494 (95% CI, 0.226 to 0.761) hovers per points from V-Month to V-Week, $p < .001$.

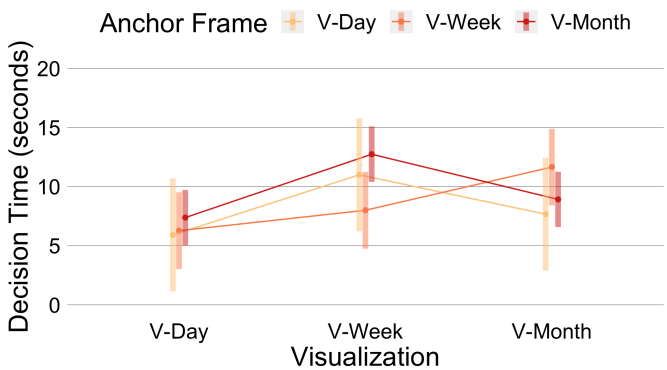


Fig. 5. Estimated marginal means of time to choose a hospital based on the anchor frame and the temporal framing of the assessed visualization.

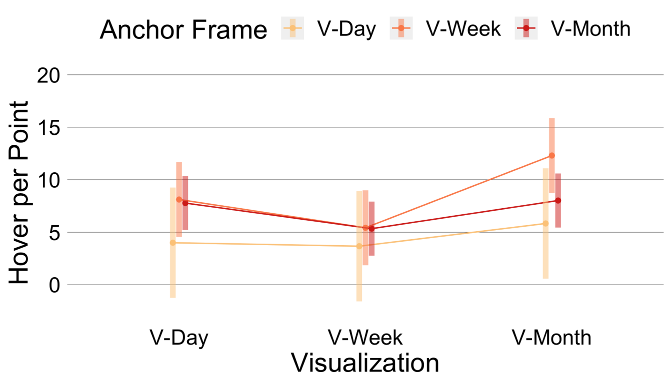


Fig. 6. Estimated marginal means of hovers per points based on the anchor frame and the temporal framing of the assessed visualization.

C. Validity of the Study

We decided to analyze whether any external artifacts affected our study. Regarding user familiarity (Figure 7), we found that most individuals familiarized themselves with line charts (4.47 ± 0.95). Since we asked participants to choose between two data trends in each time frame of the line charts, we need to check whether the decision depends on the hour, weekday, or day of the month when we conducted the test. In particular, we started by analyzing whether the hour, weekday, or day affected the anchor frame. We ran chi-square tests of independence for each time dimension (hour, weekday, day of the month). Results showed that neither the hour ($\chi^2(22) = 15.718, p = .830$), weekday ($\chi^2(10) = 6.196, p = .799$), or day of the month ($\chi^2(36) = 33.056, p = .609$) had an impact on the decision made by the participants.

Afterward, we ran one-way ANOVAs to analyze whether the hour, weekday, or day when participants conducted the study affected the time participants took to choose a hospital and decide which visualization they trusted the most when participants analyzed the visualization for the first time. Again, results showed that neither the hour ($F(5, 83) = 0.443, p = .817$), the weekday ($F(11, 77) = 1.453, p = .167$), or the day of the month ($F(18, 70) = 0.680, p = .819$) had a statically significant impact on the time taken by participants. Then, we

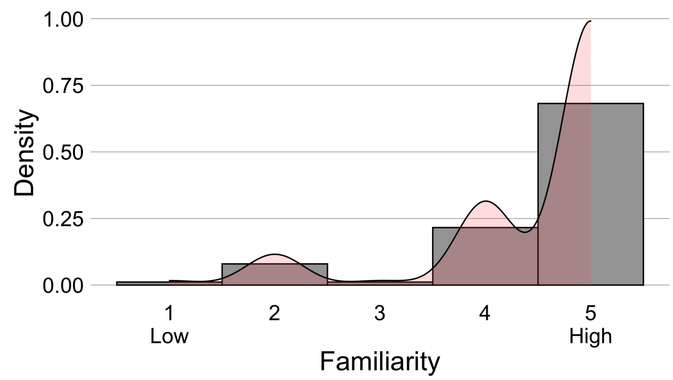


Fig. 7. Histogram with density plot of line chart familiarity.

conducted one-way ANOVAs to analyze the impact of the moment participants conducted the study on the trust perception of each of the framing values. Results showed that neither the hour, the weekday, nor the day of the month had any impact on the trust perception level for each framing visualization. Finally, we verified that the last time each participant visited a hospital did not affect the dependent variables. In particular, the distribution shows that most participants visited a hospital last year (85.39%), and similar amounts to last month (5.62%), week (3.37%), or never (4.49%). These results lead us to believe that the time the subjects conducted our study did not affect their decision-making and trust perception.

V. Discussion

Our results shed new light on the understanding of the effects of temporal framing on trust perception in the context of InfoVis with healthcare data in an emergency.

A. Answering the Research Question

Although there was a non-significant relationship between temporal framing and the trust perceived by the participants, they were more likely to attribute a lower score of perceived trust to V-Week. The high number of hover events per point and higher decision times between hospitals while interacting with V-Week corroborate these results. However, past research in Perceptual Psychology related to subitizing – "the immediate apprehension of the exact number of items in small sets" [28], [29] – suggests that individuals should find it easier to interpret visualizations with a lower number of points. In our case, we expected that individuals would have decreased decision times in V-Week since it had fewer data points and, consequently, individuals could subitize them more easily. In this light, we assume that the individuals actively spent time assessing their trust for each graph, providing some robustness to our results. Moreover, we noticed that while analyzing V-Week, people paid more attention to the peaks than the evolution of the data. It hints toward the data analysis focusing on maximum or minimum values rather than an overall appreciation of the time series.

Contrarily to this trend, results showed that, in general, participants made a decision faster when using V-Day and that they trusted more in V-Month. Regarding V-Day, the short decision time may be an influence of not aggregated (daily) *vs.* aggregated (V-Month and V-Week) values and/or that the concept of patients per hour maps directly to V-Day and, therefore, they make decisions faster. Additionally, participants may have interacted more with V-Week to know more or understand the values behind the aggregation. Contrary to what we expected, our results showed that the number of points presented in a visualization had a more predominant impact on perceived trust and user interactions than the temporal framing of the visualization. This finding is in line with Xiong et al. [10], since participants relied more on a visualization that showed a higher amount of information.

We also analyzed the self-calibrated degree of confidence in a decision by firstly asking the participants to state which temporal frame they would find more reliable in deciding between hospitals. V-Month was the most chosen level when we asked subjects which frame was more relevant to deciding between the two hospitals. Additionally, participants who initially trusted the most in that frame significantly perceived it with higher trust than the remaining framing options. In contrast, participants who initially weighed A-Week the most were more likely to change their decision about which frame weighted the most on their decision after a careful analysis of the data.

B. Additional Findings

We were able to recognize some patterns related to the analysis of time-series visualizations. In general, **participants seem to trust more in visualizations that disclosure information through more data points.** In particular, V-Month displayed 30 points, and it was the one participants trusted the most. This finding was exacerbated through the think-aloud protocol since participants mentioned that they felt more confident predicting future events when analyzing data from a higher period of time. Although V-Month has a higher time range, it is actually the same amount of information, namely the number of patients per hour with any additional aggregation information. Nevertheless, participants may wrongly perceive it as more data points. The fact that participants appeared to trust a visualization with more data points is in line with Xiong et al. [10].

Additionally, we were able to understand that **participants appeared to attribute more relevance to the maximum values when compared to the overall variation of the data when less information is displayed** such as in V-Week (7 points). We noticed for interaction data that **people interact more with a visualization that they trust less**. Researchers may leverage this relationship to adapt the framing of visualization when the system detects a large amount of interaction data since it may indicate that users are not trusting in what they are seeing.

Finally, we repeated the first component of the procedure to reevaluate the change in trust perception from the participant's perspective. As we can observe, there are no significant changes between the decisive moments. Most of the individuals kept the anchor frame after interacting with each one individually. Therefore, we conclude that asking people to rate visualizations explicitly individually does not seem to significantly influence their ratings afterward when all the visualizations are together.

C. Limitations and Future Work

Some relevant factors may explain the lack of significance observed in the effect of temporal framing on perceived trust. The assessment of perceived trust through a Likert scale may have confused participants, as they have assessed their perceived trust quite similarly in each frame. Another possible explanation may be that factual tasks may not have been enough to trigger significant trust variations. Moreover, results showed longer completion times and increased hover events performed when participants analyzed V-Week. Taking a closer look at the V-Week line chart (Figure 1b), the low amount of data points may have led participants to have an exacerbated perception of the peaks. Notice that V-Month (Figure 1c) also has the same peaks in the last seven days, yet the amount of data points reduces the area they cover, hence their reduced impact. Therefore, this design may have led to additional user interaction when participants explored this visualization, so future experiences should use randomly generated datasets. Additionally, future work should do follow-up experiments to understand whether temporal framing, the amount of data shown, or an interaction of the two drive the effect. We can consider different data encodings and a mechanism to differentiate between accidental and intentional hover events. Another limitation is that the V-Day data points are not a subset of the V-Week or V-Month charts. Consequently, the mental model varies since they have to consider new data points compared to the other line charts. These implications are crucial for the experiment and may induce bias in the results.

Future work may leverage temporal aggregation in the visualization design. A general example of temporal aggregations is the use of moving average in the finance domain, where the moving average sums up the data points of financial security over a specific time period and divides the total by the number of data points to obtain an average. In this case, the moving average is continually recalculated based on the latest price data[1]. This approach allows data to be smoother and less unpredictable [30]. Therefore, we could use this approach to study whether the moving average of patients in the different temporal frames affects trust perception.

Future work also includes analyzing perceived trust in light of individual differences such as personality. This type of analysis may explain in-depth relationships between psychological constructs and the trust level of a specific decision. In addition, leveraging a think-aloud protocol to study the use of words, such as "unsure, uncertain, maybe, perhaps", may hold a more accurate assessment of trust compared to the Likert

[1]https://corporatefinanceinstitute.com/resources/knowledge/other/moving-average/

scale assessment. In particular, these words can be indicative of uncertainty or trust, as suggested by Sacha et al. [6]. Finally, it is imperative to mention that uncertainty plays a relevant role in the trust building [6], [7], [31]. Since trust increases when users are aware of the presence of uncertainty in data [6], it is also important to represent this factor while analyzing trust perception.

VI. Conclusions

This study focused on identifying the effects of temporal framing on trust perception in the context of time-oriented visualization. Results show that temporal framing is a relevant feature of time-based visualizations since people trust more in a line chart with more data points and interact more with a line chart that they trust less. The temporal framing that subjects initially rely on the most in the decision tasks also impacts the trust perception when individuals examine visualizations one at a time. These contributions are relevant for designers of visual analytic systems, particularly when studying human decision-making supported by visualization. They provide implications to understand trust perception in a health emergency scenario and its variation inline chart techniques differing in the time-based granularity.

Acknowledgment

This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with references SFRH/BD/144798/2019 and UIDB/50021/2020.

References

[1] M. Card, *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.

[2] Y. Zhang, R. K. Bellamy, and W. A. Kellogg, "Designing information for remediating cognitive biases in decision-making," in *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 2015, pp. 2211–2220.

[3] E. Dimara, G. Bailly, A. Bezerianos, and S. Franconeri, "Mitigating the attraction effect with visualizations," *IEEE transactions on visualization and computer graphics*, vol. 25, no. 1, pp. 850–860, 2018.

[4] E. Dimara, S. Franconeri, C. Plaisant, A. Bezerianos, and P. Dragicevic, "A task-based taxonomy of cognitive biases for information visualization," *IEEE transactions on visualization and computer graphics*, vol. 26, no. 2, pp. 1413–1432, 2018.

[5] A. Kale, M. Kay, and J. Hullman, "Visual reasoning strategies for effect size judgments and decisions," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 2, pp. 272–282, 2020.

[6] D. Sacha, H. Senaratne, B. C. Kwon, G. Ellis, and D. A. Keim, "The role of uncertainty, awareness, and trust in visual analytics," *IEEE transactions on visualization and computer graphics*, vol. 22, no. 1, pp. 240–249, 2015.

[7] E. Mayr, N. Hynek, S. Salisu, and F. Windhager, "Trust in information visualization," in *EuroVis Workshop on Trustworthy Visualization (TrustVis), Robert Kosara, Kai Lawonn, Lars Linsen, and Noeska Smit (Eds.). The Eurographics Association. DOI: http://dx. doi. org/10.2312/trvis*, vol. 20191187, 2019.

[8] J. Thomas and J. Kielman, "Challenges for visual analytics," *Information Visualization*, vol. 8, no. 4, pp. 309–314, 2009.

[9] R. Salway, R. Valenzuela, J. Shoenberger, W. Mallon, and A. Viccellio, "Emergency department (ed) overcrowding: evidence-based answers to frequently asked questions," *Revista Médica Clínica Las Condes*, vol. 28, no. 2, pp. 213–219, 2017.

[10] C. Xiong, L. Padilla, K. Grayson, and S. Franconeri, "Examining the components of trust in map-based visualizations," 2019.

[11] A. K. Schnackenberg and E. C. Tomlinson, "Organizational transparency: A new perspective on managing trust in organization-stakeholder relationships," *Journal of Management*, vol. 42, no. 7, pp. 1784–1810, 2016.

[12] E. Costante, J. Den Hartog, and M. Petkovic, "On-line trust perception: What really matters," in *2011 1st Workshop on Socio-Technical Aspects in Security and Trust (STAST)*. IEEE, 2011, pp. 52–59.

[13] H.-K. Kong, Z. Liu, and K. Karahalios, "Trust and recall of information across varying degrees of title-visualization misalignment," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–13.

[14] L. Bartram, M. Correll, and M. Tory, "Untidy data: The unreasonable effectiveness of tables," *arXiv preprint arXiv:2106.15005*, 2021.

[15] A. Dasgupta, S. Burrows, K. Han, and P. J. Rasch, "Empirical analysis of the subjective impressions and objective measures of domain scientists' visual analytic judgments," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 1193–1204.

[16] K. Kelton, K. R. Fleischmann, and W. A. Wallace, "Trust in digital information," *Journal of the American Society for Information Science and Technology*, vol. 59, no. 3, pp. 363–374, 2008.

[17] A. Dasgupta, J.-Y. Lee, R. Wilson, R. A. Lafrance, N. Cramer, K. Cook, and S. Payne, "Familiarity vs trust: A comparative study of domain scientists' trust in visual analytics and conventional analysis methods," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 1, pp. 271–280, 2016.

[18] D. Sprague and M. Tory, "Exploring how and why people use visualizations in casual contexts: Modeling user goals and regulated motivations," *Information Visualization*, vol. 11, no. 2, pp. 106–123, 2012.

[19] M. M. Roghanizad and D. J. Neufeld, "Intuition, risk, and the formation of online trust," *Computers in Human Behavior*, vol. 50, pp. 489–498, 2015.

[20] W. Aigner, S. Miksch, W. Müller, H. Schumann, and C. Tominski, "Visual methods for analyzing time-oriented data," *IEEE transactions on visualization and computer graphics*, vol. 14, no. 1, pp. 47–60, 2007.

[21] M. Adnan, M. Just, and L. Baillie, "Investigating time series visualisations to improve the user experience," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, pp. 5444–5455.

[22] M. Waldner, A. Diehl, D. Gračanin, R. Splechtna, C. Delrieux, and K. Matković, "A comparison of radial and linear charts for visualizing daily patterns," *IEEE transactions on visualization and computer graphics*, vol. 26, no. 1, pp. 1033–1042, 2019.

[23] N. Oscar, S. Mejía, R. Metoyer, and K. Hooker, "Towards personalized visualization: Information granularity, situation, and personality," in *Proceedings of the 2017 Conference on Designing Interactive Systems*, 2017, pp. 811–819.

[24] W. Javed, B. McDonnel, and N. Elmqvist, "Graphical perception of multiple time series," *IEEE transactions on visualization and computer graphics*, vol. 16, no. 6, pp. 927–934, 2010.

[25] J. Heer, N. Kong, and M. Agrawala, "Sizing the horizon: the effects of chart size and layering on the graphical perception of time series visualizations," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2009, pp. 1303–1312.

[26] D. V. de Alwis and C. H. Kon, "A new way to use the ishihara test," *Journal of neurology*, vol. 239, no. 8, pp. 451–454, 1992.

[27] J. Cohen, *Statistical power analysis for the behavioral sciences*. Academic press, 2013.

[28] H. Railo, M. Koivisto, A. Revonsuo, and M. M. Hannula, "The role of attention in subitizing," *Cognition*, vol. 107, no. 1, pp. 82–104, 2008.

[29] M. Piazza, A. Fumarola, A. Chinello, and D. Melcher, "Subitizing reflects visuo-spatial object individuation capacity," *Cognition*, vol. 121, no. 1, pp. 147–153, 2011.

[30] A. Raudys, V. Lenčiauskas, and E. Malčius, "Moving averages for financial data smoothing," in *International Conference on Information and Software Technologies*. Springer, 2013, pp. 34–45.

[31] J. Zhou, S. Luo, and F. Chen, "Effects of personality traits on user trust in human–machine collaborations," *Journal on Multimodal User Interfaces*, pp. 1–14, 2020.