Int. J. Human–Computer Studies 191 (2024) 103343

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Visualisations with semantic icons: Assessing engagement with distracting elements

Muna Alebri [a,*], Enrico Costanza [a], Georgia Panagiotidou [b], Duncan P. Brumby [a], Fatima Althani [c], Riccardo Bovo [d]

[a] University College London, Interaction Centre, 66-72 Gower St, London, WC1E 6EA, London, United Kingdom
[b] Department of Informatics, King's College London, London, United Kingdom
[c] School of Electronic Engineering and Computer Science, Queen Mary University of London, London, United Kingdom
[d] Imperial College London, Computer Science Department, London, United Kingdom

ARTICLE INFO

Keywords:
Embellished visualisations
Visualisations for the public
Engagement

ABSTRACT

As visualisations reach a broad range of audiences, designing visualisations that attract and engage becomes more critical. Prior work suggests that semantic icons entice and immerse the reader; however, little is known about their impact with informational tasks and when the viewer's attention is divided because of a distracting element. To address this gap, we first explored a variety of semantic icons with various visualisation attributes. The findings of this exploration shaped the design of our primary comparative online user studies, where participants saw a target visualisation with a distracting visualisation on a web page and were asked to extract insights. Their engagement was measured through three dependent variables: (1) visual attention, (2) effort to write insights, and (3) self-reported engagement. In Study 1, we discovered that visualisations with semantic icons were consistently perceived to be more engaging than the plain version. However, we found no differences in visual attention and effort between the two versions. Thus, we ran Study 2 using visualisations with more salient semantic icons to achieve maximum contrast. The results were consistent with our first Study. Furthermore, we found that semantic icons elevated engagement with visualisations depicting less interesting and engaging topics from the participant's perspective. We extended prior work by demonstrating the semantic value after performing an informational task (extracting insights) and reflecting on the visualisation, besides its value to the first impression. Our findings may be helpful to visualisation designers and storytellers keen on designing engaging visualisations with limited resources. We also contribute reflections on engagement measurements with visualisations and provide future directions.

1. Introduction

In an era dominated by visual content and big data, visualisations have become widely consumed by a broad audience, whether through news outlets, blogs or social media. Such visualisations reach audiences that may not have a strong intrinsic motivation to explore them. Visualisations for the public may carry vital information that shapes our decisions and futures, and overlooking them may lead to a missed knowledge opportunity. Understanding the unique characteristics of these visualisations has been on the visualisation community agenda for over a decade. Prior work (Pousman et al., 2007; Lee et al., 2020) emphasised the importance of moving away from the narrow definition of traditional visualisation audiences where everyone 'needs' to look at visualisation and extract a value. Sprague and Tory (2012) illustrate that people apply a cost-benefit model to use visualisations. People invest effort and time into these visualisations not only to gain knowledge but also to achieve a pleasing experience.

Designers historically employed images, illustrations, and icons in visualisations to engage their audience. Researchers have seen growing interest in evaluating the effectiveness of these design choices (e.g., Bateman et al., 2010; Borkin et al., 2013). Still, scholars are at odds in describing these elements and their value. For instance, some use 'Chart Junk' (e.g., Tufte, 1983; Ajani et al., 2021) while others advocate for less negative concatenation, such as embellishments and enhancements (e.g., Akbaba et al., 2021; Peña et al., 2020). In this work, we use the term 'embellishment' and define it as *visual elements (e.g., icons, pictographs, backgrounds, shapes) that make a visualisation deviate from a standard visualisation*. Fig. 1 provides an illustrative example of embellished and standard visualisations.

* Corresponding author.
E-mail addresses: muna.alebri.19@ucl.ac.uk (M. Alebri), e.costanza@ucl.ac.uk (E. Costanza), georgia.panagiotidou@kcl.ac.uk (G. Panagiotidou), d.brumpy@ucl.ac.uk (D.P. Brumby), f.althani@qmul.ac.uk (F. Althani), r.bovo19@imperial.ac.uk (R. Bovo).

https://doi.org/10.1016/j.ijhcs.2024.103343
Received 5 March 2024; Received in revised form 8 July 2024; Accepted 20 July 2024
Available online 23 July 2024
1071-5819/

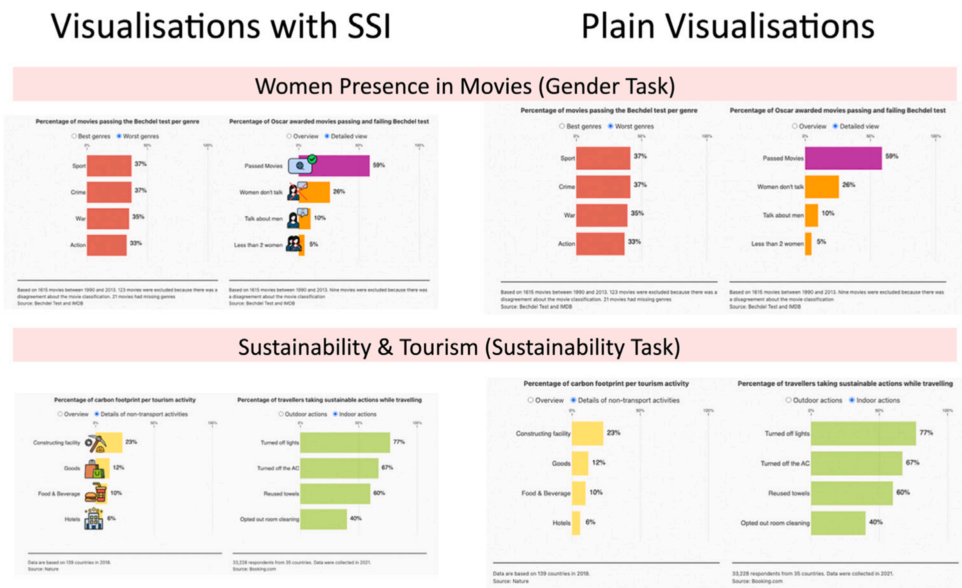


Fig. 1. Web page examples from Study 1. The top segment shows the Gender topic under both conditions (with and without semantic icons). The bottom segment shows the Sustainability topic under both conditions.

Prior work that explored the impact of embellishments (e.g., stacked pictographs, icons) on engagement provides evidence of their positive effects (Haroz et al., 2015; Andry et al., 2021; Alebri et al., 2023; Burns et al., 2021). The definition of engagement varies between these contributions. For the scope of this paper, 'Engagement' is defined as *the user's investment into exploring a visualisation* (Boy et al., 2015). We translate this investment into the time and effort spent writing insights from the visualisations. Moreover, most existing literature evaluated embellished visualisations with a free exploration task. We evaluate embellished visualisations with an informational task (extracting insights) to understand how the aspect of effort and time impact engagement. Previous work (e.g., Cleveland and McGill, 1984; Saket et al., 2019) pointed out the vital role of tasks in evaluation studies and how performance differs for different task types. Furthermore, engagement with visualisations has been primarily examined in isolation, excluding distractions that may help understand the saliency of the embellished visualisation and how the viewer's attention is distributed under cognitive load. Distractions have been leveraged in the literature to understand conscious attention to visualisations (e.g., Haroz et al., 2015), visual phenomena (e.g., Simons and Chabris, 1999) and reflect complex visual environments (e.g., Navalpakkam and Churchill, 2012). We employ distractors to investigate whether the positive impact of embellishments exists in situations where there might be a competing demand for the audience's attention (unconsciously).

The area of embellishments is vast, and many types are yet to be explored. In this work, we explore 'Semantic icons (SI)': *icons and pictographs that provide a visual representation of the data in the chart* – for instance, using a flag next to a country's name or a money icon next to an income label (e.g., see Fig. 1). Despite being widely used in practice, less empirical evidence exists around their effectiveness on engagement. Shi et al. (2022) demonstrated through analysing a large set of pictorial visualisations from the wild (published visualisations in the media and other sources) that associating an icon to the meaning of a category is the most common data binding strategy (32% of the dataset). Additionally, such embellishments have been widely integrated into online visualisation authoring tools (e.g., Flourish[1] and Datawrapper[2]). Considering the gaps identified, we form the following research question: *How do SI affect engagement with visualisations when there are distractions while extracting insights?*

To answer our research question, we first explored a variety of SI with various visualisation attributes. Based on the results of this exploration, we designed and carried out two comparative online studies. In Study 1, participants were presented with two web pages containing two visualisations- a target and a distracting one. Participants saw the target with or without SI and were asked to extract insights, and later, they filled out a survey about their engagement with the target visualisation. We found that participants rated the SI version as more engaging than the plain version. Nevertheless, we did not find evidence of a difference in (a) visual attention: time spent on the target visualisations and (b) cognitive effort: the insights from that visualisation between the two versions. We suspected the results were due to the SI being too subtle. Thus, we replicated the study (Study 2) with visualisations from the wild containing salient SI (prominent, large-sized) and other embellishment types for maximum contrast.

The two studies revealed that SI, simple or more salient, enhanced perceived engagement even with distractions and after putting effort into extracting insights. Furthermore, our results suggest that aesthetic appeal was an important dimension that motivated participants to report the SI version as more engaging. In addition, visualisations with less interesting and engaging topics benefited from SI. Despite the positive outcomes, we could not reproduce these findings using our visual attention and cognitive effort metrics. Overall, our findings make the following contributions: (1) empirical evidence showing that SI are valuable after performing an informational task, such as extracting insights and reflecting on the visualisation, besides being valuable for the first impression. This contribution informs visualisation designers keen on engaging their readers but with limited time to design embellishments. (2) Lessons learned on using distractions, visual attention, and cognitive effort metrics in measuring engagement based on our primary and exploratory studies, which may guide future researchers in the field.

2. Background

2.1. Visualisations for the public

Visual representations are one of the approaches used as a solution to simplify large datasets and complex analyses. Nevertheless, until recently, the visualisation community operated under the assumption that

[1] https://flourish.studio/.
[2] https://www.datawrapper.de/.

all consumers of visualisations belonged to a single population, namely, experts in domains and visualisations (Lee et al., 2020; Hullman et al., 2011). This presumption has implications, including the belief that visualisations are predominantly utilised for professional purposes, with individuals dedicating extended periods to their examination (Pousman et al., 2007). Therefore, visualisations have been evaluated with respect to usability goals (e.g., error and speed). A review of visualisation evaluation studies by Choi et al. (2023) revealed that cognitive experiments (e.g., engagement, memorability) were rare in empirical investigations where they represent 21% of the total experiments. Similarly, Borgo et al. (2018) found that 60% of crowdsourcing visualisation evaluations used error and speed, while other measures, such as enjoyment and engagement, received less attention. We contribute to this space and focus on *engagement*, which is discussed in detail in Section 2.3.

A body of research acknowledged that applying the same assumptions used with traditional visualisations initially designed for experts with visualisations for the public is impossible. Peck et al. (2019) interviewed visitors of farmers' markets in Pennsylvania to understand how education and economic status contribute to what attracts them to visualisations. They found that people were attracted to simple and familiar visualisation types (e.g., bar and line charts) and visualisations with topics that spoke to them or their circles. He et al. (2023) explored the lived experiences of the general public with energy visualisations and suggested that receptivity to visualisations is not unified; some responses were data enthusiastic while others were information avoidant. Through focus groups with diverse interests, backgrounds, and skills, Kennedy et al. (2016) revealed that topic, time, emotions, source, beliefs and opinions, skills and confidence with visualisations are factors that contribute to the viewer's engagement with visualisations. These contributions are valuable as they shed light on a population group often overlooked. Yet, further investigation is needed beyond the data aspect of these visualisations (e.g., design aspect, context).

2.2. Semantic icons

Relatively limited literature focused mainly on SI, although embellishments have been debated in the community for a long time. Associating icons with data is a widespread technique in practice. Setlur and Mackinlay (2014) indicated that 30.52% of Tableau public galleries contained categorical data utilised icons. Employing icons may be easy and does not require specific design skills. Tools such as icon8[3] and flaticon[4] allow automatic icon generation in various styles using simple keywords.

Past research showed that icons, pictograms, and human-recognisable objects enhanced memorability. For instance, Borkin et al. (2013) found that pictograms made visualisations as memorable as images. Extending their work, Borkin et al. (2016) asked participants to recall the content of the visualisations, and they found that data or message redundancy, such as using icons and pictographs, enhanced recall and understanding. Arunkumar et al. (2023) found consistency in how viewers internalised visualisations; in particular, visualisations with icons and pictographs were rated more as image-like rather than information-like and were seen to benefit the emotional impact and aesthetic appeal.

On the other hand, evidence suggests caution in employing SI. For instance, when icons or pictographs replace text labels, they affect memorability negatively (Haroz et al., 2015). Similarly, Peña et al. (2020) evaluated various embellishments, including backgrounds and stacked pictographs and found that memorability was enhanced only when the embellishment was relevant to the data. Borgo et al. (2012) also found that embellishments (various types, including SI) slowed the visual search but assisted concept grasping. Skau and Kosara (2017) investigated pictorial bar charts in a crowdsourced study and found them to have a limited impact on accuracy if used within bar bounds.

Different strategies of employing SI have been evaluated on their effectiveness in engaging the reader. For instance, Haroz et al. (2015) compared stacked pictographs with bar charts and text by showing participants these elements in a grid and asking them to freely explore and click on the thumbnail for a full view. They found that participants' initial attention was directed more frequently to stacked pictographs than bars and text. Another study that employed free exploration task is by Andry et al. (2021) where they explored a variety of embellished visualisations, including SI. They interviewed media professionals and asked them to look at visualisations collected from the media and rate them based on beauty, interest, clarity, ease of understanding, and preference. They found that visualisations with icons and pictographs were particularly preferred as they immersed them in the visualisation topic.

On the contrary, a recent study by Burns et al. (2021) evaluated engagement with icon arrays using informational tasks where participants were asked to answer several questions targeting different levels of engagement. They did not find a significant difference between icon arrays and traditional visualisations in attracting participants to write lengthy responses or put more effort into composing them. However, participants reported that icon arrays made them envision the data behind the visualisation. These results emphasise the importance of exploring different embellishment strategies and expanding our knowledge about engagement with informational tasks.

2.3. Assessing engagement with visualisations

Seeking a cohesive definition of engagement proves to be complicated. Doherty and Doherty (2018) reviewed how engagement is defined within the Human–computer interaction community and found that only 65% of 351 reviewed papers explicitly defined engagement, leaving the community with an ambiguous understanding of engagement. Hung and Parsons (2017) found a similar pattern in the Information Visualisation research and suggested considering the following factors in the definition: (1) intention: the viewer has the initial motivation and curiosity; (2) autonomy: the viewer decides to continue the interaction; (3) purpose: not based on a utilitarian goal; (4) time: spending more than a few seconds, and (5) outcome: extracting more than a data point. Yet, considering all these factors in comparative user evaluations is challenging since they require some control.

The difficulty of reaching a unified definition of engagement stems from the nature of the concept of being multifaceted, making it challenging to capture all at once. Some researchers focused on understanding when defining engagement, such as Mahyar et al. (2015). They described engagement as *the level of user involvement in understanding the visualisation*, which ranges from viewing to deriving a decision. Insights captured is another measure used as an indirect proxy of engagement. For instance, Vande Moere et al. (2012) and Wood et al. (2012) considered the insights' quantity and quality.

Others captured engagement from an interaction perspective. For instance, Hung and Parsons (2017) used mouse movement to indicate people's attention to a particular visualisation. Mouse data has been utilised extensively in the literature to demonstrate attention and, thus, engagement. Mouse data complement gaze data and provide a cheaper and less intrusive option. Prior work (Chen et al., 2001; Huang et al., 2011) demonstrated a high correlation between eye gaze and mouse movement. Mouse data has been used to measure engagement while performing web search tasks (e.g., Navalpakkam and Churchill, 2012) and news reading (e.g., Lagun and Lalmas, 2016). Some researchers (Haroz et al., 2015) used the proportion of participants who looked at a visualisation over time to measure the viewer's attention indirectly. Others utilised time, number of visits, and mouse interactions (e.g., clicks and hover) to measure engagement with storytelling

[3] https://icons8.com/.
[4] https://www.flaticon.com/.

visualisations (Boy et al., 2015). Time has also been used to measure engagement with visualisations beyond the digital format. Saket et al. (2016) and Wang et al. (2019b) displayed visualisations as posters and tracked how long participants paused to read the visualisation.

Further, a body of work investigated engagement from the user's subjective perspective using surveys to evaluate the different components of engagement (e.g., Andry et al., 2021; Wang et al., 2019b; Amini et al., 2018). Surveys can be distributed on a large scale and are easy to employ; therefore, it is the most used method in Human–Computer Interaction to measure engagement (Doherty and Doherty, 2018). Inspired by the contributions mentioned above, we measure engagement through the following elements: (1) visual attention, (2) effort to write insights, and (3) perceived engagement (self-reported).

2.4. Distractions in visual environments

To grasp people's interactions with embellished visualisations, one must consider the impact of distractions and how they shape the reader's perception and behaviour. Yet, very few visualisation studies utilised distractions in their approach. One of these rare studies is by Borgo et al. (2012) where they evaluated embellishments' impact on visual perception using a dual-task methodology. Participants were asked to perform a primary task (e.g., extracting a value) while being interrupted with a secondary task (clicking on fruit names) to mimic divided attention. Another example is by Haroz et al. (2015) where they employed a blurred 3 × 3 grid with stacked pictographs, bar charts, and text and participants were instructed to click on the element they wanted to inspect further to remove the blurring. The blurring aspect distorts the visualisations, resulting in a conscious and slower viewing process. Distractions have also been utilised in evaluating web content noticeability. For instance, Navalpakkam and Churchill (2012) used advertisements and their position as distractions to determine how content noticeability changes from a simple web page layout. Simons and Chabris (1999) demonstrated how distracting elements could attract attention and influence awareness of salient features in a visual environment. Leveraging distractions may be beneficial in evaluating the visibility of embellishments and relevance to our complex visual environments, which are often bombarded with many elements.

3. Exploratory study: impact of semantic icons

Our primary research question (RQ) is to understand how visualisations with SI impact engagement when viewers encounter them in complex settings while extracting insights. The complex setting refers to distractions in the visual environment to understand whether SI make the visualisation more salient and motivates participants to allocate their attention. Furthermore, we aim to explore the feasibility of employing voluntary participation to reflect some aspects of the visualisations people encounter daily. Next, we describe our approach to address the RQ.

3.1. Method

3.1.1. Participants

Having received approval to conduct the study from the UCL ethics review board, we recruited 94 participants via mailing lists, a subject pool, and social media. We excluded participants who did not complete the visualisation tasks ($n = 7$) and had technical issues ($n = 1$). Therefore, the following analysis is based on the remaining 86 participants, of which demographic data were available for 83. Among our participants providing these data, 43 were women, 36 were men, one was non-binary, and three preferred not to disclose. Participants' ages ranged between 18 and 55 years old ($M = 27.9$, $SD = 7.4$). Participants' education levels were as follows: 40 participants had an undergraduate degree, 26 had a graduate degree, 10 had a doctorate, and seven had other qualifications. Three participants indicated that they have a form of colour deficiency. We decided to keep the data from these participants as the colours in our visualisations were accessible using Adobe colour simulator (Adobe, 2021). Participants' familiarity with visualisations was medium-to-high in a 5-point Likert scale (1 5) ($M = 3.81$, $SD = 0.72$) adapted from Pandey et al. (2015). Participants voluntarily participated in the study. To encourage participation and motivate participants to provide several insights, we awarded a prize of £50 to the top three participants who provided the highest number of meaningful insights (see Section 3.3.2 for a detailed description of insight classification).

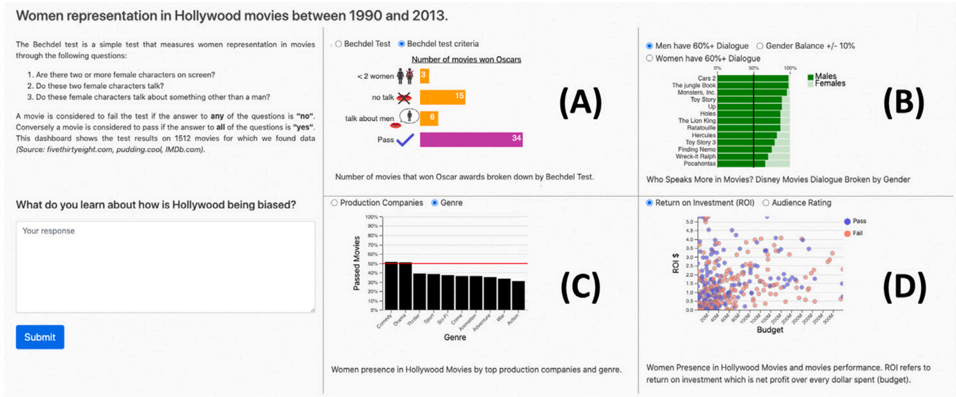
3.1.2. Study design

For our study, participants saw two web pages on two topics (for simplicity, we refer to them as 'Gender' and 'Success'), each containing four visualisations (A, B, C, D). We used counterbalancing for the web page order when assigning participants to the study condition. Each participant saw the web page either with all plain (P) visualisations (e.g., [A_P, B_P, C_P, D_P]) or with one visualisation that has SI and three plain visualisations (e.g., [A_{SI}, B_P, C_P, D_P]) (see Fig. 2 for an example). We used randomisation to decide which visualisation to add the SI to. Our experimental approach to display a target visualisation (the one with SI) with distraction is motivated by the following: (1) to reflect the cluttered visual environments people encounter visualisations at in different realms, (2) distractions may simulate a cognitive load requiring participants to allocate their attention to a stimulus that has an advantage over others, and (3) gain insights about how visualisation salience is maintained even with distractions. Moreover, investigating visualisation designs using this approach may apply to visualisation thumbnail designs in the news, where specific designs may entice viewers more than others to explore the full article further.

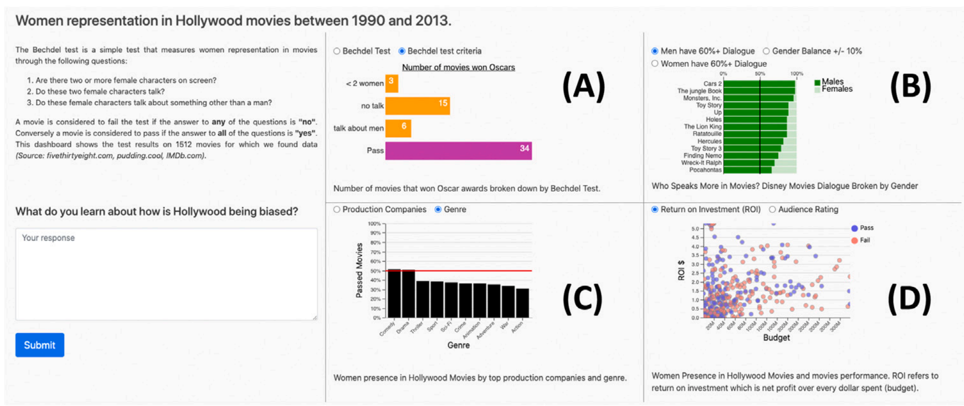
We randomly assigned participants to the study conditions: 66 participants saw each web page (Gender and Success) with a visualisation that had SI, and 20 saw the plain version of these web pages. Since participation was challenging (as voluntary) and conditions were randomised, we had varying numbers of participants for each visualisation. For the SI condition with the Gender web page, 14 participants saw visualisation A with SI, 15 saw B with SI, 17 saw C with SI, and 20 saw D with SI. For the Success web page, 15 saw visualisation A with SI, 15 saw B with SI, 17 saw C with SI, and 19 saw D with SI. We later compare each visualisation that has SI with its plain version (e.g., in the Gender web page, we compare A_{SI} with A_P) to test for differences in our dependent variables.

Engagement is a complex, multidimensional concept, which motivates measuring it with multiple dependent variables (DV):

1. **Visual attention**, which we measured through time ratio that was calculated by dividing the total time a mouse cursor was within a target visualisation area (e.g., visualisation A) by the total time spent on all visualisations (A, B, C, D). The time ratio was considered from when the page was rendered to when the insights were submitted. We used time ratio instead of absolute time because we did not restrict participants on time and to overcome individual differences.
2. **Cognitive effort**. Specifically, we measured (1) insights ratio, the number of insights captured from a visualisation divided by the total number of insights extracted for the task, and (2) effort level, each visualisation insight was classified as low or high based on the participant's effort of extracting and interpreting the data (more details including example are provided in Section 3.3.2).
3. **Perceived engagement: self-reported**, which we captured through a survey (refer to Section 3.1.4 for details).



(i) An example of the Gender web page where visualisation A is embellished



(ii) An example of the Gender web page where visualisation A is plain

Fig. 2. Web page examples: each participant saw one version.

3.1.3. Materials

We chose two datasets about movies to create the visualisations. These datasets enabled us to recruit participants from the general public rather than requiring specific expertise. The first dataset (A), for the Gender web page, included information about the presence of women in Hollywood movies using the Bechdel test. We expected this dataset to be controversial: some people would strongly oppose discrimination against women. The second dataset (B), for the Success web page, included information about the most profitable Hollywood movies.

To create dataset A visualisations, we replicated three available visualisations with modification from Anderson and Daniels (2016), Hickey (2014) and used the official IMDb database (IMDb.com, 2020) to create the fourth visualisation. Dataset B was neutral: We expected fewer people to have strong opinions about this dataset. To create dataset B visualisations, we used the official IMDb database. We produced two interactive web pages for each dataset using the D3 library in Javascript.[5] We provide the code to reproduce the visualisations, data, and analysis in Alebri et al. (2024). Each web page contained four visualisations and a short description of the dataset. The web page was displayed to participants on a single screen and required minimum scrolling for a screen of 1024*768 pixels. All the visualisations within each web page were consistent in size and font style. We added captions under each visualisation to explain it. All visualisations had two interaction techniques: filtering and details on demand.

We collaborated with a graphic designer in creating and placing the icons. We also searched for standard practices for representing each icon idea and followed that with pilots. We aimed to design simple, universally understood icons that display a single idea as much as possible. To include the icons, occasionally, we had to shift labels slightly; otherwise, we maintained the same style as the plain version. Figs. 3 and 4 show how we employed SI in our visualisations. Standard practices inspired the position of the SI in visualisations. We also wanted to explore how SI could be used with various data types. We had visualisations with categorical variables (e.g., Fig. 3-A) and non-categorical variables (e.g., Fig. 4-A).

We employed an 'informational task' inviting participants to extract insights from the visualisations. We followed an open-ended approach and started the questions with: "What do you learn about. . . ?". We asked participants about the visualisation theme to mitigate feeling overwhelmed. In doing so, we guided their exploration to a specific goal. We decided not to restrict participants to a time limit to ensure they explore the visualisation at their own pace.

3.1.4. Procedure

Participants were required to use a desktop or a laptop with at least 1024*768 pixels to ensure the visualisation quality. We verified their participation eligibility by running a script that checks screen resolutions. We developed the online user study using the Django framework[6] and hosted it locally on a web server at our university. Upon consenting and reading the instructions, which included an example of the task, participants saw the first web page and its task, followed by the User Engagement Scale (UES) from O'Brien and Toms

[5] https://d3js.org/.

[6] https://www.djangoproject.com/.

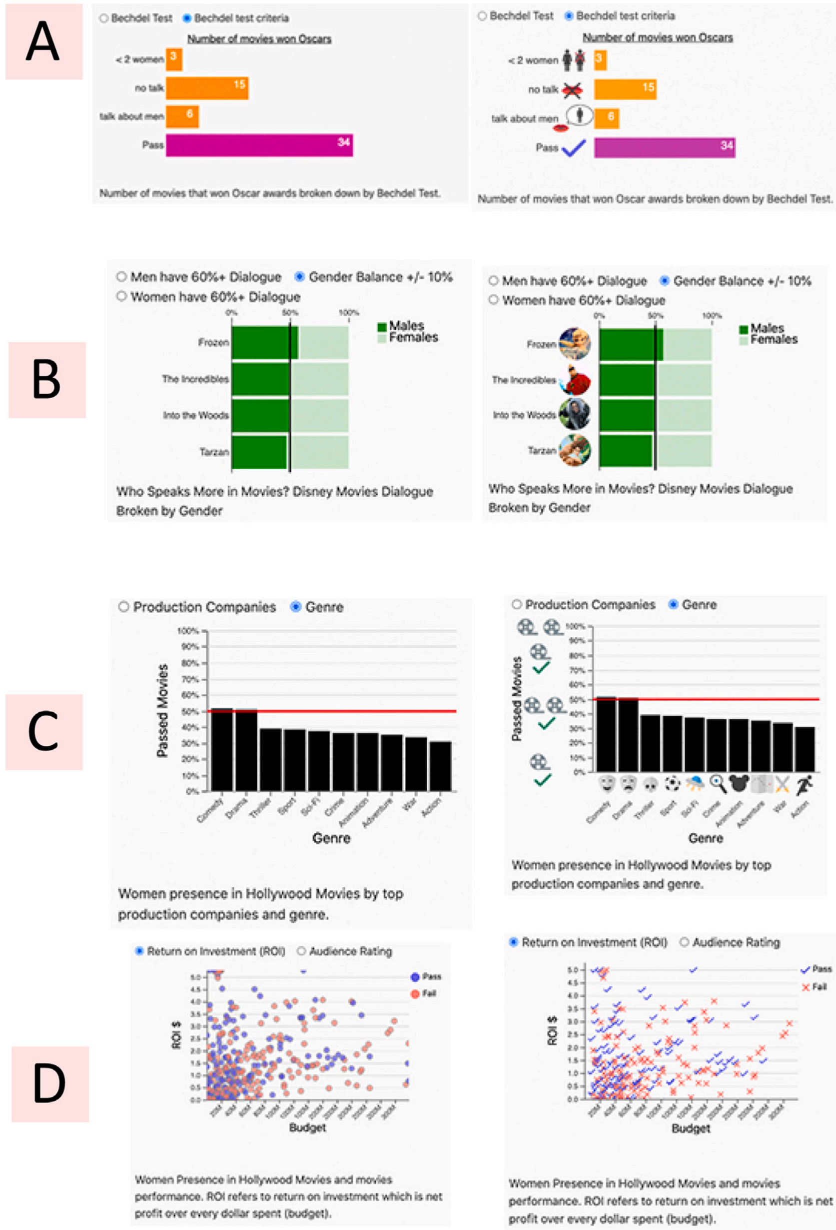


Fig. 3. Examples of the visualisations with SI (right) and their plain version (left) of the Gender web page.

(2010). The survey included 12 questions and assessed four dimensions: focused attention, perceived usability, aesthetic appeal, and reward factor. We adopted the short survey version from O'Brien et al. (2018) as the original survey, and the informational task (insight extraction) might exhaust participants. The survey has been validated with at least 40 published studies (O'Brien et al., 2018), including data visualisation, such as Wang et al. (2019b). Finally, participants were invited to write some feedback about their experience with the study, their chart familiarity, and demographics. Fig. 5 provides an overview of the study procedure.

3.2. Hypotheses

Our main RQ is how SI impact engagement with visualisations when attention is divided and directed to extract insights. We developed the following hypotheses to address our RQ:

- H1: We expect the visual attention measured as the ratio of time spent on visualisation with SI in a web page to be higher than with the plain version.
- H2.a: We expect the cognitive effort measured as the insights ratio (number of insights captured from the target visualisation divided by the total insights) to be higher for participants who saw the SI version than those who saw the plain version.
- H2.b: We expect participants to extract more low-level and high-level insights from visualisations with SI than the plain version.
- H3: We expect participants to rate web pages with a visualisation that contains SI as more engaging than the web page with the plain version.

3.3. Results

Before analysing participants' engagement within each visualisation, we provide an overview of participants' completion time (duration

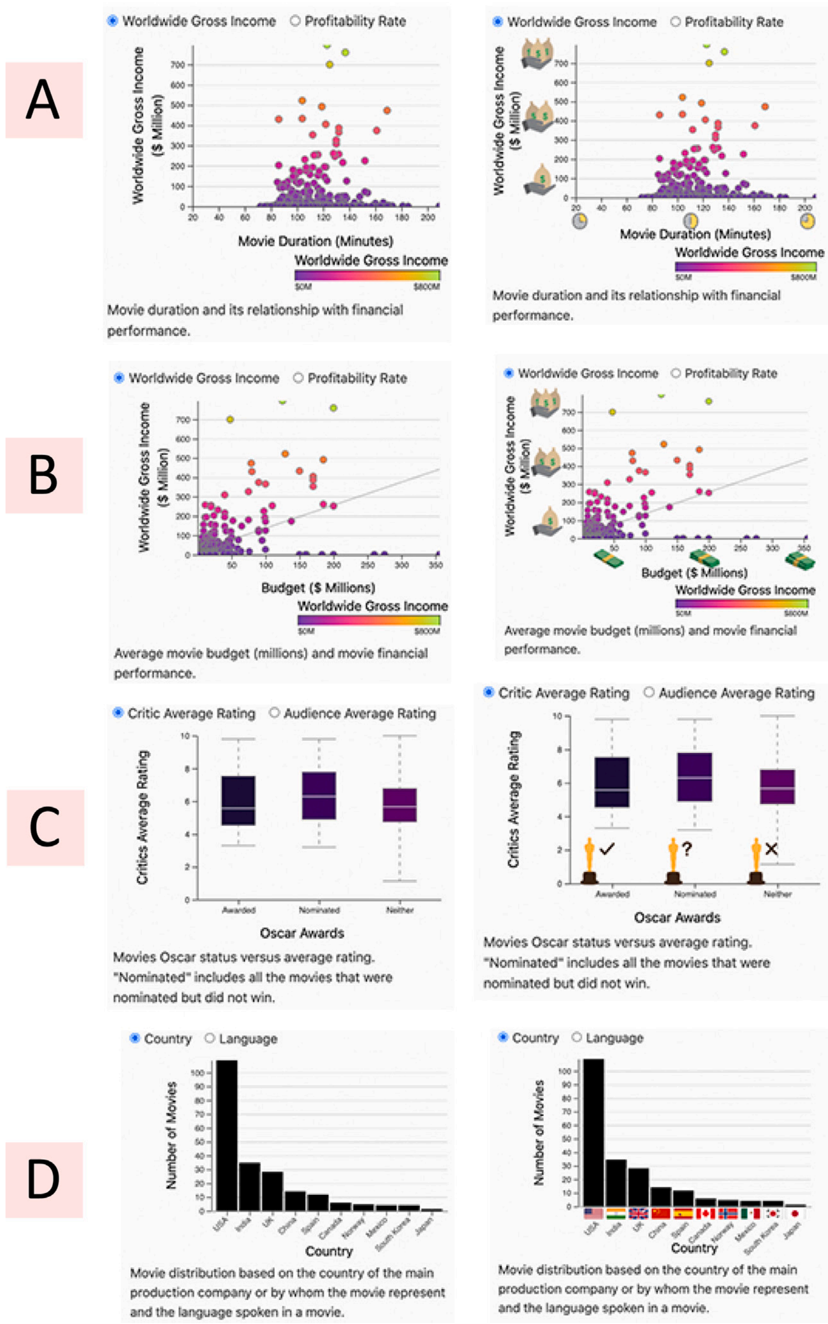


Fig. 4. Examples of the visualisations with SI (right) and their plain version (left) of the Success web page.

from the first render to submitting a response) for each web page (Gender and Success). On average, participants spent 17 min and 71 s (SD = 1646) on the Gender web page, while they spent 29 min and 71 s (SD = 8712) on the Success web page. Next, we report the results based on the DV of each visualisation and compare the SI version with its counter (e.g., A_{SI} versus A_P).

3.3.1. *Visual attention: Mouse movement*

We calculated the total time a participant's cursor was within a visualisation to evaluate participants' visual attention to the visualisations. We recorded mouse events based on mouse enter and leave times. Next, we calculated the time ratio by dividing a cursor's time in a particular visualisation by the total time a cursor was within the four visualisations. Since we had a non-equal sample size with a normal distribution, We performed a Welch t-test on the effect of SI on visual attention based on the time ratio spent in visualisation areas. Although H1 was partially supported with visualisation B in the Gender web page, the lack of evidence of an effect and a trend for the rest of the visualisations makes it difficult to reach a conclusion. As shown in Fig. 6-A.1, we observed in the Gender web page that participants spent significantly longer on the SI version of **visualisation B** (M = 0.36, SD = 0.16) compared to its plain version (M = 0.26, SD = .11), ($t(23.8)$ = −2.09, p = .048), CI95% [−1.423, −0.022]). Further, Cohen's effect size value revealed a large effect (d = .7). We did not observe the same effect with **visualisation A** ($t(31.9)$ = 0.358, p = .723), **visualisation C** ($t(33.8)$ = 0.350, p = .729), and **visualisation D** ($t(35.3)$ = 0.306, p = .761).

Regarding the Success web page, as shown in Fig. 6-A.2, we observed no significant difference in time ratio between the SI and the plain version of the four visualisations: **visualisation A** ($t(30.2)$ = 1.05,

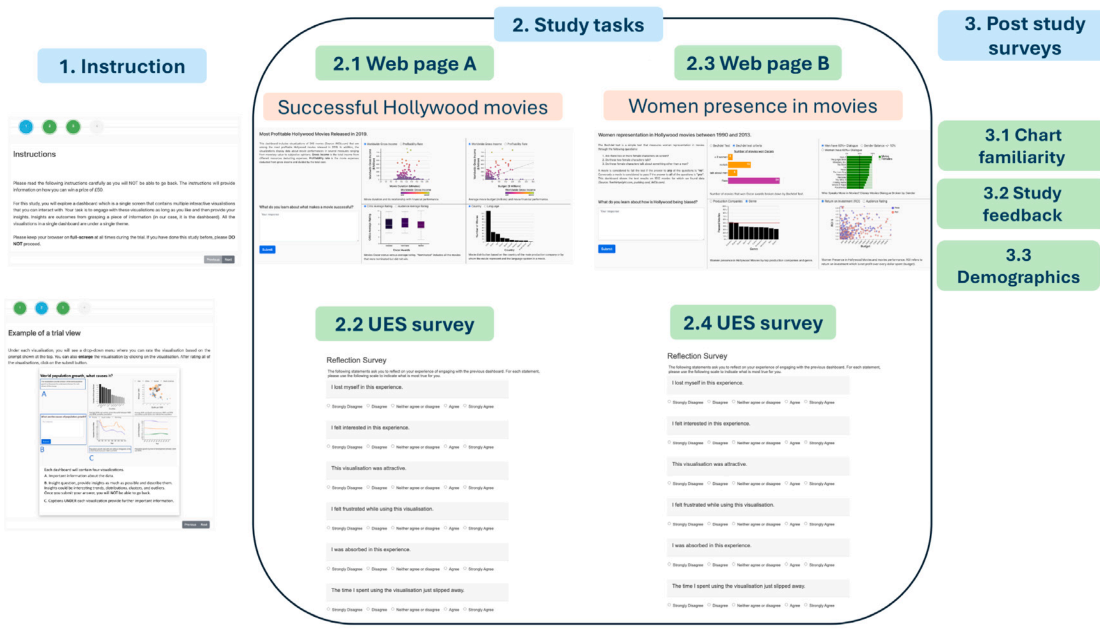


Fig. 5. Illustrative diagram of the study procedure.

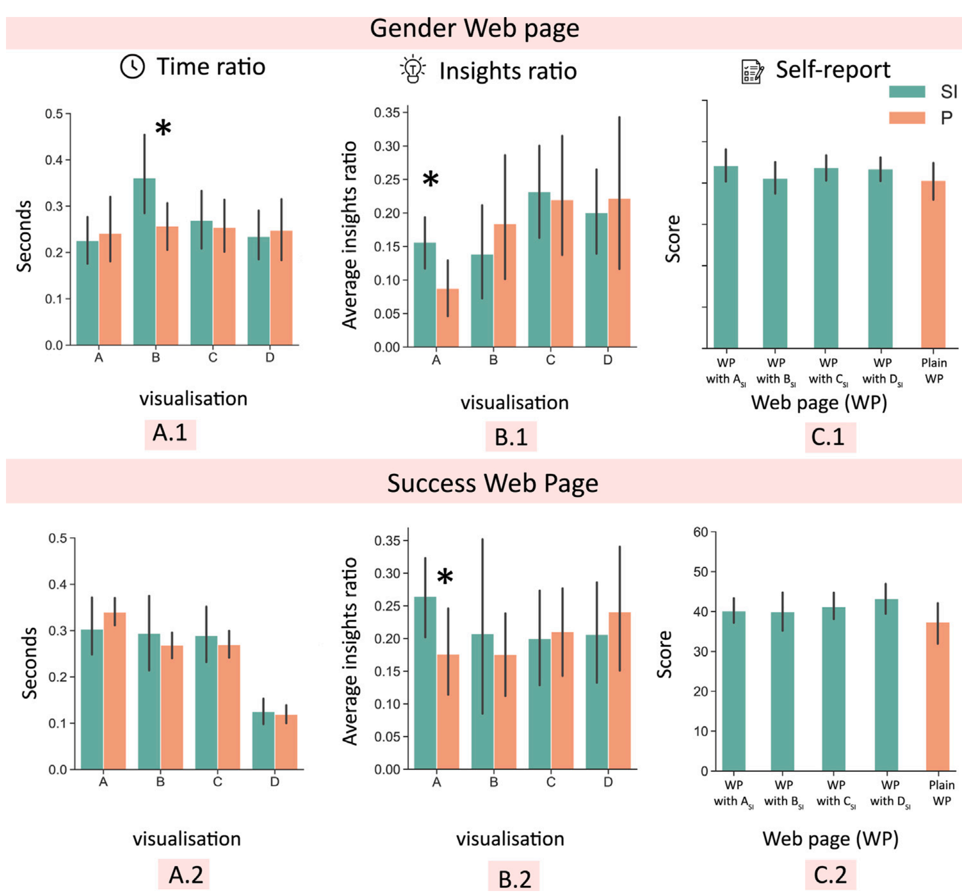


Fig. 6. Engagement metric results organised by web page. A.1 and A.2 show the average time ratio (total time a cursor was within a visualisation over the total time on the four visualisations) for SI and plain (P) visualisations. B.1 and B.2 display the average insights ratio: the number of insights captured from each visualisation divided by the total insights. C.1 and C.2 display the total engagement score based on self-reported data for each web page. Significant differences are noted with $*$ and based on $p < .05$. Error bars represent 95% confidence intervals.

p = .301), **visualisation B** ($t(20.54) = -1$, p = .329), **visualisation C** ($t(34.98) = 0.01$, p = .991), and **visualisation D** ($t(36.57) = -0.66$, p = .513).

3.3.2. Cognitive effort: Extracting insights

We asked participants to share what they learned from the visualisation. Participants wrote a total of 1067 insights. We analysed the insight quantity **(insights ratio)** and quality **(effort level)**.

Insights ratio. We calculated the **insights ratio** for each participant to indicate allocating more attention and effort to a visualisation. To calculate the insights ratio, we divided the number of insights extracted from a visualisation by the total number of insights extracted during a task. We decided to use the insights ratio instead of the absolute number of insights to account for individual differences, as participants were not restricted on time and the number of insights extracted. We also analyse the absolute number of insights captured from the target visualisation in Appendix A. Next, we describe our analysis process to calculate the insights ratio.

First, we categorised the insights into their related visualisation (A, B, C, D). Next, We disaggregated insights whenever participants described more than two variables. Whenever organising an insight under a single visualisation was impossible, we classified it under *general insights*. These insights were either about the meaning of the visualisations (n = 44) or usability insights (n = 77) inspired by Vande Moere et al.'s (2012) and Chen et al.'s (2009) insight classifications. *Meaning insights* did not fit a single low-level task (e.g., retrieve a value, trend, distribution) or a compound task. For instance, insights that included opinions, interpretations and personal experiences (e.g., P77 wrote under the Gender web page: *"In other words, women can be there simply for eye-candy purposes and to have someone for the hero to save"*). *Usability insights* were about the data itself or the visualisation design, such as the presence and absence of data, font size, colours, or interactivity (e.g., P89 wrote about the Success web page: *"overall difficult to identify any causal effects from the given data"*). Furthermore, we excluded meaningless responses (n = 4) from the analysis, such as responses restating the web page summary and responses challenging to understand due to bad grammar (e.g., P44 wrote about the Gender web page: *"I learned about the Bechdel test and that people can apparently always find something to complain about"*. We used the Mann–Whitney U test in the analysis.

Insights-ratio. Although H2.a was partially supported with visualisation A in the Gender web page and visualisation A in the Success web page, the lack of evidence of an effect and trend for the rest of the visualisations makes it difficult to reach a conclusion. For the Gender web page, we found a significant difference in insights ratio between SI and plain version for **visualisation A** (U = 206, p = .02, CI95%[0.107,0.720]) of a medium effect ($Cohen's\ d = 0.5$), suggesting that when the visualisation had SI, participants were more inclined to extract insights from the target visualisation relative to total insights compared to the plain version as shown in Fig. 6-B.1. We did not find evidence of a difference between the SI and the plain version of **visualisation B** (U = 142, p = .78), **visualisation C** (U = 182, p = .73), and **visualisation D** (U = 212, p = .75).

For the Success web page (see Fig. 6-B.2), we found a difference in the insights ratio between SI and plain version for **visualisation A** (U = 211, p = 04, CI95%[0.036, 0.675]), ($Cohen's\ d = 0.4$). Specifically, participants who saw the SI version had a higher insight ratio than those who saw the plain version. However, we did not find a difference in the insights ratio between the SI and the plain version for **visualisation B** (U = 140, p = .73), **visualisation C** (U = 162, p = .82), and **visualisation D** (U = 178, p = .73).

Effort-level. Regarding the effort level, each insight extracted from the target visualisation was classified into one of the following: *low-level effort* or *high-level effort*. An insight is classified under *low-level effort* if the insight includes data from the visualisation without interpretation. For example, under the Success web page, the insight *"Most films cluster around the 80–140 min duration and towards the much lower end on budget"* from P84 was classified as low-level effort insight. An insight was classified as a *high-level effort* insight if the participant added their interpretation and expression or predicted future values. For instance, we classified P55 insight under the Gender web page: *"Additionally, most movies that have won Oscars that fail the test fail because women do not talk. While this would be purely speculation, this seems to indicate that Hollywood includes women in films so they are 'represented' but they are ultimately extras and could be removed without any effect on the film"* as a high-level effort insight.

H2.b was not supported, as we found conflicting results with the low-level insights, and we found no evidence of a difference for the high-level insights. Regarding the **low-level insights**, the Mann–Whitney U test under the Gender web page revealed that there is a significant difference in the number of low-level insights written between the SI version and the plain version (U = 212, p = .007, CI95% [0.167, 0.748]) for **visualisation A** with a medium effect ($Cohen'sd$ = 0.5). In particular, participants wrote more low-level insights when they saw the SI version (M = 1.21, SD = 0.70) than the plain version (M = 0.5, SD = 0.70). However, we did not find the same effect with **visualisation B** (U = 157, p = .83), **visualisation C** (U = 195, p = .44), **and visualisation D** (U = 219, p = .59).

The test for the Success web page revealed an opposite effect with **visualisation B**. We found that participants wrote significantly more low-level insights with the plain version (M = 1.15, SD = 0.93) compared with the SI version (M = 0.53, SD = 0.83) (U = 95, p = .05, CI95% [−0.653, 0.004]) and the effect was medium ($Cohen'sd$ = 0.5). The wide confidence interval underscores the uncertainty of the effect observed with visualisation B. We did not find an effect for **visualisation A** (U = 177, p = .34), **visualisation C** (U = 174, p = .93), and **visualisation D** (U = 200, p = .79).

Regarding **high-level effort insights** under the Gender web page, the test suggests that there are no differences in the insights extracted between the SI version and the plain version for **visualisation A** (U = 135, p = .81), **visualisation B** (U = 129, p = .25), **visualisation C** (U = 172, p = .97), and **visualisation D** (U = 149, p = .60). We found a similar result for the Success web page. We did not find evidence of a difference in insights between the SI version and the plain version for **visualisation A** (U = 137, p = .45), **visualisation B** (U = 164, p = .48), **visualisation C** (U = 136, p = .31), and **visualisation D** (U = 192, p = .97).

3.3.3. Perceived engagement: Self-reported

We calculated the total engagement score for each web page based on the responses to 12 questions. We compared the engagement score of a plain web page with a web page that has a visualisation with SI. For instance, we compared the total engagement score of a plain web page with the total engagement score of a web page with visualisation A_{SI}. We used the Mann–Whitney U test to evaluate the effect of SI on perceived engagement. We used a non-parametric test because data from Likert-scales is 'ordinal' (not 'interval' or 'ratio') (Jamieson, 2004).

We did not find evidence to support H3, which states that a web page with SI will be rated higher on engagement than web pages with all plain visualisations. For the Gender web page, we did not find a significant difference in the total engagement score between the plain web page version and the web page with **visualisation $\mathbf{A_{SI}}$** (U = 158, p = .539) as shown in Fig. 6-C.1. Similarly, we did not find evidence of a difference for web pages with **visualisation $\mathbf{B_{SI}}$** (U = 142, p = .789), or **visualisation $\mathbf{C_{SI}}$** (U = 200, p = .367), or **visualisation $\mathbf{D_{SI}}$**. Similarly, for the Success web page as shown in Fig. 6-C.2, we found no significant difference between the plain web page version and the web page with **visualisation $\mathbf{A_{SI}}$** (U = 162, p = .713), **visualisation $\mathbf{B_{SI}}$** (U = 167, p = .582), **visualisation $\mathbf{C_{SI}}$** (U = 201, p = .360), and **visualisation $\mathbf{D_{SI}}$** (U = 240, p = 0.168).

Table 1
Quantitative analysis overview of the exploratory study based on the visualisation version (SI, plain) and dependent variables.

Factor	Visualisation	Dependent variables		
		Visual attention (time-ratio)	Cognitive effort (insights-ratio)	Perceived engagement (self-reported)
Version	A_{Gender}		[a]	
	B_{Gender}	[a]		
	C_{Gender}			
	D_{Gender}			
Version	$A_{Success}$		[a]	
	$B_{Success}$			
	$C_{Success}$			
	$D_{Success}$			

[a] Significant difference is based on $\alpha = .05$.

3.4. Post-study feedback

After completing the main study tasks, participants were asked to share their feedback. Some participants described the experience as *interesting* ($n = 18$) while a few described it as *enjoyable* ($n = 6$) and *fun* ($n = 4$). Very few participants described the visualisations as *aesthetically pleasing* ($n = 3$); although they saw the SI version, they did not mention it. Some participants expressed *feeling overwhelmed* ($n = 4$) because of the number of visualisations and the data density. For instance, P25 commented: *"Although I thought that having all figures at once was useful to combine all information together, it was overwhelming when I first looked at them"*. This also may be linked to some participants reporting feeling *tired* ($n = 3$). For instance, P69 said: *"Too much information in one dashboard does not help convey a message. Instead it can get tiring and repetitive"*. Some participants compared the two pages (Gender and Success), where there was more preference for the Gender web page ($n = 6$) compared to the Success web page ($n = 1$).

4. Discussion

As the summary in Table 1 shows, we observed that SI attracted attention for one out of eight visualisations (visualisation A in the Gender web page, see Section 3.3.1) and that SI prompted writing more insights for two out of eight visualisations (visualisation A in both web pages, refer to Section 3.3.2) while observing no effect on the rest. These results prevent us from making generalisable recommendations and provide practical guidelines. Upon reflecting on our study approach and findings, we identified the following factors that need to be considered when investigating engagement with SI visualisations:

Visualisation attributes. The visualisation type (e.g., bar chart, scatter plot), number of data points, and colour are likely to interfere with the SI in relation to engagement. For instance, while we observed significant differences in two bar charts (A, B) on the Gender web page (refer to Sections 3.3.1 and 3.3.2), similar effects were not seen with other bar charts (i.e., visualisation C on the Gender web page and visualisation D on the success web page). We speculate that the number of data points and colour may have interacted with SI. Similarly, it was noted that visualisation A on the Success web page (scatter plot) encouraged writing more insights (see Section 3.3.2); on the other hand, we did not find the same effect with visualisation C on the Gender web page (scatter plot). Despite visualisation C in the Gender web page being colourful, substituting the visual mark with icons may have cluttered the visualisation, particularly given its high number of data points instead of adding SI as labels.

Semantic icons. The SI employed in the exploratory study may vary in detail, prominence, size, and position. For instance, the SI in Fig. 3-visualisation B had more details than those in visualisation-C.

Participant variability. The data variability may reflect our sample's diversity and the fact that they were recruited from various venues (social media, mailing lists). Additionally, it may reflect that participants were not restricted by time and number of insights to extract. A more constrained lab or online study may mitigate the variability. However, exploring voluntary recruitment was useful to reflect on the challenges and complexity of people's interactions with visualisations. Recruiting voluntary participants proved challenging and time-consuming, potentially due to the length of the task.

Number of distracting elements. Some participants expressed in the post-study feedback feeling overwhelmed and fatigued because of the number of visualisations on each web page in the study (see Section 3.4). We hypothesise that this sentiment may have contributed to the lack of observed differences in perceived engagement between web pages with and without SI (see Section 3.3.3), as participants were asked to assess the entire web page rather than individual visualisations.

A more controlled study may address the factors discussed above to understand the usefulness of SI in engaging people with visualisations, which we aim to achieve in Study 1.

5. Study 1: Semantic icons in horizontal bar charts

Study 1 aimed to understand whether we can derive generalisable findings about the usefulness of SI engaging readers when distracting elements are in view. The exploratory study revealed that SI interactions with visualisations are more complicated than initially anticipated, necessitating further investigations with a simplified study design and control. Therefore, for Study 1, while we maintain our approach to evaluating engagement with distracting elements, we simplify the task by presenting participants with a target visualisation alongside a distracting one. Additionally, we focus solely on horizontal bar charts, as two out of three visualisations (visualisations A and B on the Gender web page) showed that SI had a positive impact on engagement (refer to Sections 3.3.1 and 3.3.2). Furthermore, we control the topic, number of data points, and colour. We moved to an online crowdsourcing platform, Prolific,[7] to facilitate recruitment and mitigate variability. Crowdsourcing has been utilised previously to investigate visualisations for the general public (e.g., Alebri et al., 2023; Burns et al., 2021). Financial incentives for participation have also been employed previously in investigating engagement with embellished visualisations (e.g., Haroz et al., 2015). Further details about the study method are provided in the following section.

[7] https://prolific.co/.

5.1. Method

5.1.1. Participants

We recruited 58 participants (M_{Age} = 27.33, SD_{Age} = 7.43, 29 female) via Prolific. Our sample size was determined based on a power analysis that used the effect size of the visual attention metric from the exploratory study. We used Prolific built-in screening, requiring participants to be fluent in English and not colourblind. Participant's education was as follows: one participant had a doctorate, 15 had a graduate degree, 17 had an undergraduate degree, and 25 had other qualifications. Participant's chart familiarity was medium-to-high (1–5) (M = 3.95, SD = 0.62), which was similar to the participants in the exploratory study (see Section 3.1.1). In terms of skills with visualisations and design, 19% of our participants stated considering design as a hobby, 9% considered design as a profession, 7% reported creating visualisations very frequently, and 7% reported visiting visualisation forums (e.g., Statisa, r/dataisbeautiful) very often. We decided to consider all participants as a variety of visualisation and design skills would be a good representation of the general public. Participants were paid £2.25 based on an estimated study duration of 15 min and £9 hourly rate (based on UK minimum pay Government, 2022 at the time of conducting the study).

5.1.2. Study design

We followed a similar study design to the exploratory study except that on each web page, participants saw two visualisations: *a target visualisation and a distracting visualisation* (see Fig. 1). The target is the visualisation to which we apply our experimental approach (i.e., SI), and the distracting is the visualisation we keep fixed across conditions. The web page order and the visualisation position within a page were randomised. Participants were asked to extract insights from the two visualisations.

Based on the lessons learned from the exploratory study, we modified our DV as follows: (1) We added an analysis of the initial visual attention, inspired by Haroz et al. (2015), in particular, the time spent on a visualisation area within the first minute of counter divided by time spent on all visualisations. This is instead of only looking at attention allocation to visualisation from the time the page is rendered until the insights are submitted, i.e., the approach we followed in the exploratory study (see Section 3.1.2). Exploring initial attention would provide a more nuanced understanding of the elements that capture initial attention before becoming more familiar with the material. It may also enable us to isolate the viewer's attention before participants distribute their attention more equally to extract insights. To deploy this change, some modification was made to the code that tracks mouse movement, where each mouse event was recorded with a timestamp instead of storing the total time spent on an area, as we did in the exploratory study. (2) We limit the perceived engagement (self-reported) to the target visualisation rather than the entire web page as we did in the exploratory study.

We kept hypotheses (H1, H2.a, H.2,b) as stated in Section 3.2 but used our modified definition of our DV. We modified H3 as follows: *We expect participants to rate visualisation that contains SI as more engaging than the plain version*. This is because we did not observe an effect on the perceived engagement (self-reported) in the exploratory study (see Section 3.3.3). We speculate that because the dashboards were dense with information and SI may have been overshadowed upon reflection on their experience.

5.1.3. Visualisation design

For the design of the visualisations, we used two topics: (1) Women's representation in movies (for simplicity, it will be referred to as 'Gender') and (2) sustainability and tourism (will be referred to as 'Sustainability'). We used the same Gender visualisation (A) from the exploratory study as evidence showed that SI encouraged putting more effort (refer to Section 3.3.2). Both topics were inspired by real-world visualisations (see Hickey, 2014; N, 2023; Liu, 2023; International, 2018). We also chose datasets with categorical data that can be represented with SI. As shown in Fig. 1, all visualisations were colourful (bars themselves), and we maintained the same colour scheme for each visualisation. All the visualisations were horizontal bar charts to ensure simplicity and familiarity with the chart type. We also ensured that all the visualisations had the same number of data points to isolate the SI effect.

The SI were displayed in the exact location on each visualisation and adjacent to the text with semantic meaning. We aimed to design visualisations that reflect the type of visualisations found in the wild. Through an analysis of 37 visualisations collected over eight weeks (February–April 2022) from UK news outlets (BBC, Daily Mail), non-profit organisations (e.g., UNDRR), and visualisation forums (Visual Capitalist, Statista), we found that 65% of bar charts had the SI adjacent to the textual label. To ensure consistency of the icon design, including colour and expression level, we hired a graphic designer. Designing the icons was a thorough and iterative process where we constantly looked at visualisation examples and icon generation tools. Similar to the exploratory study, the visualisations were interactive. The title for each visualisation was placed on the top, while the source and any information related to the data were placed at the bottom of each visualisation. We provide the data for generating the visualisations, the icons, the data collected, and analysis in the supplementary material (Alebri et al., 2024).

5.1.4. Procedure

We followed a similar procedure used in the exploratory study (refer to Section 3.1.4) as shown in Fig. 5. In the Engagement survey, we provided participants with a screenshot of the visualisation while completing the survey to remind them of the visualisation. Next, participants were asked about their interest in the visualisation topics (i.e., 'Women representation in movies' and 'Sustainability and tourism') on a 5-point Likert scale. We added this question to understand the results further as several scholars revealed the vital role of the topic on engagement with visualisations (e.g., Alebri et al., 2023; Peck et al., 2019; Kennedy et al., 2016). Finally, to provide context about the participants, we asked them to complete a survey about their skills with visualisations and design.

5.2. Results

As we have reduced the number of visualisations and data points in each web page compared to the exploratory study, participants spent less time completing each task on Study 1. On average, participants spent 5 min and 65 s ($SD = 298$ s) on the Gender task while they spent 4 min and 60 s on the Sustainability task ($SD = 276$ s). In the following sections, we describe the results in relation to each DV.

5.2.1. Visual attention: Mouse movement

In Study 1, we considered initial attention (i.e., time spent on the visualisation area within the first minute of the encounter divided by time spent on all visualisations). We also report in Appendix B the visual attention during the entire task time to contrast with the exploratory study (see Section 3.3.1). We reached a similar result; therefore, we consider initial visual attention only. A factorial analysis has been used to investigate the impact of the topic (Gender and Sustainability), visualisation version (SI, plain), and their interaction on our DV. A factorial analysis was not suitable in the exploratory analysis (refer to Section 3) as there was less control over the topic, visualisation type, and number of data points.

Our results suggest no evidence to support H1 about SI advantage in attracting visual attention. The Aligned Rank Transform (ART) method was used as the Shapiro–Wilk test of residuals showed that the data deviates from normality (W = 0.97, p = .01). As shown in Fig. 7, the results suggest no main effect of the topic ($F(1, 56) = 1.07$, $p = .31$), no main effect of the visualisation version ($F(1, 56) = 0.26$, $p = .61$) and no interaction effect ($F(1, 56) = 0.14$, $p = .71$).

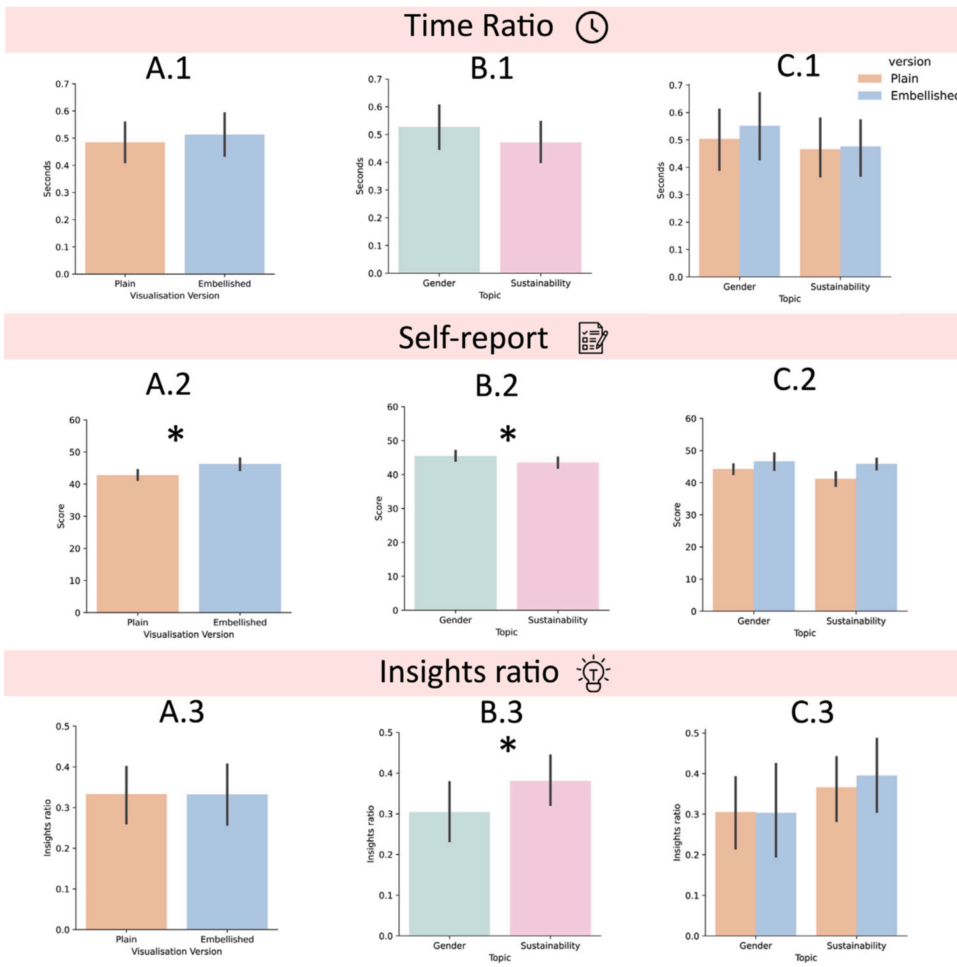


Fig. 7. Results of engagement measures: time ratio (A.1, B.1, C.1), total engagement score (A.2, B.2, C.2), and insights ratio (A.3, B.3, C.3). The results are organised by effects, left-to-right: visualisation version, topic, and interaction effect. Error bars are based on 95% confidence intervals. We measured significance based on $\alpha = .05$.

5.2.2. Perceived engagement: Self-reported

The total engagement score was calculated for each participant based on the UES scale, which aggregates 12 questions. We found evidence to support H3, which states that the visualisations with SI would be rated more engaging than their plain version. A Shapiro–Wilk test of residuals showed that the data deviates from normality ($W = 0.96$, $p = .002$), so the ART model was used for analysis. As shown in Fig. 7-A.2, there was a main effect of the visualisation version ($F(1,56) = 9.60$, $p = .003$). Specifically, participants rated the SI version ($M = 46.24, SD = 5.67$) more engaging than the plain version ($M = 42.76, SD = 5.34$). The test also reveals a main effect of the topic ($F(1,56) = 9.21$, $p = .004$). Participants gave the Gender visualisation a higher engagement score ($M = 45.47$) than the sustainability visualisation ($M = 43.53$). The test suggests no interaction between the visualisation version and topic ($F(1,56) = 1.09$, $p = .30$).

Further, the engagement scale's four dimensions (aesthetic appeal, focused attention, perceived usability, reward factor) were analysed based on O'Brien et al. guidelines (O'Brien et al., 2018). Each dimension engagement score was calculated through the answers to three questions. The analysis helps in understanding what made these visualisations to be perceived as more engaging. In summary, we found visualisations with SI were perceived as more aesthetically pleasing and rewarding than the plain version. We also found that the visualisation topic was important for perceived focused attention. Next, we report these results in more detail.

First, for the **aesthetic appeal** dimension, the ART model reveals a main effect of the visualisation version ($F(1,56) = 15.56$, $p = .0002$). In particular, participants rated visualisations with SI to have a higher aesthetic appeal ($M = 12.67$) than plain visualisations ($M = 11.03$). On the other hand, the test suggests no main effect of the topic ($F(1,56) = 0.31$, $p = .08$) and no interaction effect ($F(1,56) = 0.50$, $p = .48$). Second, for the **focused attention** dimension, the test suggests no main effect of the visualisation version ($F(1,56) = 1.90$, $p = .17$). But the topic had a main effect ($F(1,56) = 6.18$, $p = .02$). In particular, participants perceived that the Gender topic ($M = 9.33$) attracted their attention more than the Sustainability topic ($M = 8.69$). The test suggests no interaction effect between the topic and the visualisation version ($F(1,56) = 0.23$, $p = .63$).

Third, for the **perceived usability** dimension, there was no main effect of the visualisation version ($F(1,56) = 1.90$, $p = .17$), no main effect of the topic ($F(1,56) = 3.28$, $p = ..08$), and no interaction effect ($F(1,56) = 0.68$, $p = .41$). Fourth, for the **reward factor** dimension, there was a main effect of the visualisation version ($F(1,56) = 3.96$, $p = .05$). In particular, participants rated visualisations with SI more rewarding ($M = 11.97$) than plain visualisations ($M = 11.26$). On the other hand, there was no main effect of the topic ($F(1,56) = 2.46$, $p = .12$) and no interaction effect ($F(1,56) = 2.84$, $p = .10$).

5.2.3. Cognitive effort: Extracting insights

Next, we report our analysis of the insights ratio and effort level (See Section 3.3.2 for a detailed description of the metrics). We also analyse the absolute number of insights in Appendix C, which reach the same conclusions as the insights ratio analysis.

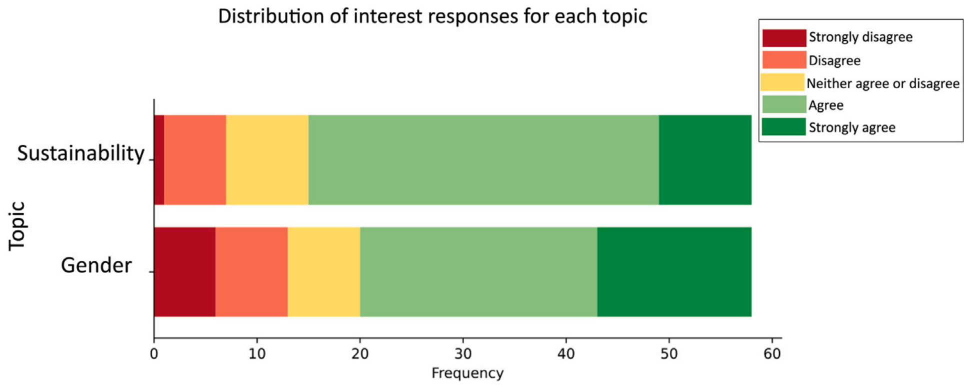


Fig. 8. Distribution of interest responses based on participants' agreement to the following statement: "I am interested in the topic of $< topic >$".

Insights ratio. We followed a similar insight analysis approach to the exploratory study (see Section 3.3.2). We did not find evidence to support H2.a, which expects the insight ratio to be higher for participants who saw the SI version compared to those who saw the plain version. As shown in Fig. 7-A.3, the ART model suggests no main effect of the visualisation version ($F(1,56) = 0.63$, p = .43). However, the topic had a main effect ($F(1,56) = 4.59$, $p = .04$) on the insights ratio. In particular, the insight ratio was higher for the Sustainability topic ($M = 0.36$, $SD = 0.23$) compared to the Gender topic ($M = 0.30$, $SD = 0.28$). The test revealed no interaction effect between the visualisation version and topic ($F(1,56) = 0.19$, $p = 0.66$).

Effort-level. We did not find evidence to support H2.b, which evaluates whether the presence of SI encouraged writing low and high-level insights. The ART model suggests that under **low-level insights**, there was no main effect of the topic ($F(1,56) = 0.52$, $p = .47$), no main effect of visualisation version ($F(1,56) = 0.78$, $p = .38$), and no interaction effect ($F(1,56) = 0.02$, $p = .89$).

Regarding **high-level insights**, the test suggests no main effect of visualisation version ($F(1,56) = 3.00$, $p = .09$). However, the topic had a main effect ($F(1,56) = 6.44$, $p = .01$). In particular, participants wrote more expressive and interpretative insights (see Section 3.3.2 for detailed definition and classification) under the Sustainability topic ($M = 0.29$, $SD = 0.75$) compared to the Gender topic ($M = 0.07$, $SD = 0.32$). The test revealed no interaction effect between the visualisation version and topic ($F(1,56) = 1.27$, $p = .26$).

5.2.4. Topic interest

Although interest was not in the initial research question, it may help understand the results. Fig. 8 illustrates the distribution of participants' interests across the topics. The ART method was used as the Shapiro–Wilk test of residuals showed that the data deviates from normality ($W = 0.91$, $p < .001$). The test suggests no main effect of the visualisation version ($F(1,56) = 2.37$, $p = .13$), no main effect of the topic ($F(1,56) = 0.08$, $p = .78$), and no interaction effect ($F(1,56) = 0.31$, $p = .58$) on interest level.

5.3. Discussion

As the results summary in Table 2 show, there are two key takeaways from Study 1: First, SI made the visualisation be perceived as more engaging than their plain version, and this was because of the aesthetic appeal and the reward factor. Section 7 discusses how these results may be helpful in practice. Second, the topic was also influential in perceived engagement and insights. We also observed no differences in visual attention between visualisation versions and topics. We speculate that the subtle differences between the SI and plain versions may have contributed to the lack of an effect on visual attention. For instance, if the icons were removed, the target and the distracting visualisations would be similar in style. Using a visualisation with salient SI would maximise the contrast with a plain version and help validate the usefulness of time and insight ratio as proxies of engagement when distracting visualisations accompany the target visualisation. We, therefore, decided to run a follow-up online user study to validate our speculations, which will be discussed in the following section.

6. Study 2: Salient semantic icons

This study aims to investigate the validity of our visual attention and cognitive effort measures in capturing the difference in engagement between different visualisation versions (SI, P). To approach our goal, we employ visualisations that would achieve maximum contrast. We expected that visualisations with multiple embellishment types that contain prominent and salient SI may be suitable for this purpose. Instead of crafting visualisations for the user study, we decided to use visualisations from the wild to reflect current design practices better. Specifically, we utilised visualisations from UK news outlets (The BBC and The Daily Mail) (see Fig. 9). We employed these visualisations based on the strong indications from our prior work (Alebri et al., 2023) that viewers found these visualisations enticing and attractive. We were also encouraged to use these visualisations as they were less subtle than those in Study 1. We should be able to detect if there is an effect on visual attention and cognitive effort.

6.1. Method

We followed the same approach used in Study 1 (see Section 5). The only difference was that in Study 2, we used static visualisations from the wild in our target visualisations, which we describe in Section 6.1.2 compared to interactive visualisations in Study 1.

6.1.1. Participants

We recruited 58 participants ($M_{Age} = 38.57$, $SD_{Age} = 12.57$, 29 female) via Prolific. Our participants had medium-to-high chart familiarity (1 – 5) ($M = 3.7$, $SD = 0.69$), similar to the participants in Study 1 (see Section 3.1.1) and the exploratory study (see Section 5.1.1). Regarding participants' skills in visualisation and design, 10% of our participants stated considering design as a hobby, 3% considered design as a profession, 9% created visualisations very frequently, and 3% visited visualisation forums very often. Participants' education was as follows: two participants had a doctorate, 12 had a graduate degree, 22 had an undergraduate degree, and 22 had other qualifications.

6.1.2. Visualisation design

We had two web page topics: Food prices and Space rockets. The target visualisations were taken as they are from their sources. We created the plain version of the target by removing pictorials, arrows

Table 2
Quantitative analysis overview of Study 1 based on study factors and dependent variables.

Factor	Dependent variables			
	Visual attention (time-ratio)	Cognitive effort (insights-ratio)	Perceived engagement (self-reported)	Interest
Version (SI, Plain)			[a]	
Topic		[a]	[a]	
Interaction (Topic X Version)				

[a] Significant difference is based on $\alpha = .05$.

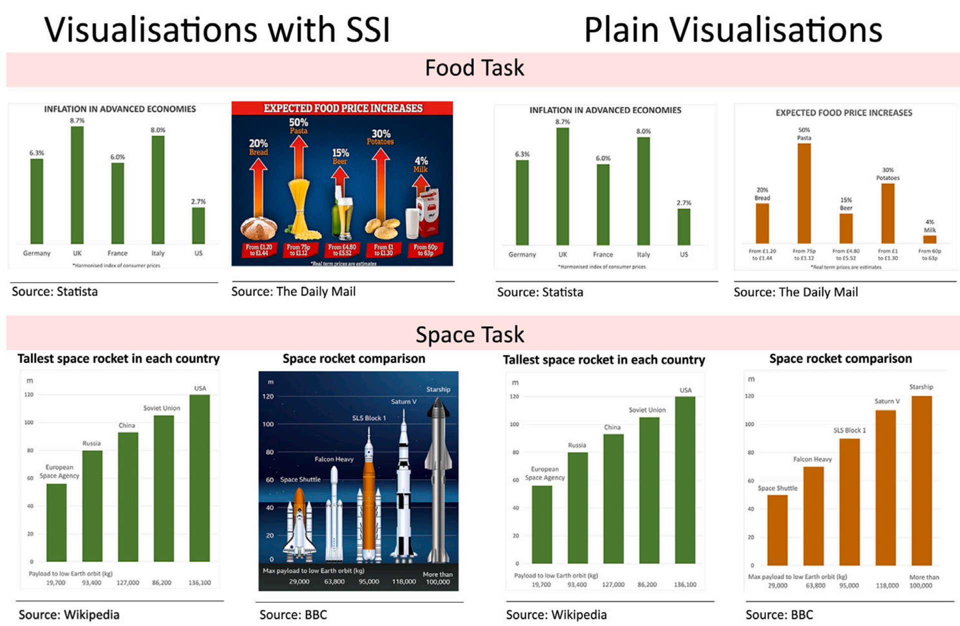


Fig. 9. Web page examples from Study 2 that employ salient SI. The top part shows the two conditions of the Food topic, while the bottom part shows the two conditions of the Space topic. Each participant saw only one of the conditions for each web page.

and backgrounds. We also streamlined the font colour and size for maximum contrast. The distracting visualisations were also inspired by real-world visualisations (see Launch vehicle, 2024; Race, 2023). All plain visualisations were created using Microsoft Excel. Fig. 9 shows an example of each web page in both versions. Because the target visualisations were taken from the media, we had limited control over their design. Refer to the supplementary material (Alebri et al., 2024) for the visualisations, data, and analysis.

6.2. Results

We used the same approach from Section 5.2 to analyse the data of the three DV as a proxy of engagement: visual attention, cognitive effort, and perceived engagement (self-reported). As the visualisations in this study were static, participants spent less time on each web page than in Study 1. On average, participants spent 2 min and 13 s ($SD = 92.4sec$) on the Food web page and spent 2 min and 55 s ($SD = 105sec$) on the Space web page.

6.2.1. Visual attention: Mouse movement

We did not find evidence to support H1, which states that visualisations with SI would attract initial attention compared to the plain version. Initial visual attention was measured as follows: the time spent on the target visualisation divided by the total time spent on the target and the distracting visualisations within the first minute of the encounter. We used the ART method as the Shapiro–Wilk test of residuals showed that the data deviates from normality ($W = 0.82$, $p < .0001$). As shown in Fig. 10, the test suggests no main effect of the visualisation version ($F(1, 56) = 0.58$, $p = .44$), no main effect of the topic ($F(1, 56) = 0.93$, $p = .34$), and no interaction effect ($F(1, 56) = 0.12$, $p = .73$) on visualisation time ratio. We also report visual attention results during the entire task time in Appendix D. Both analysis approaches reached the same finding: there was no evidence of a difference in attention between the SI and plain versions.

6.2.2. Perceived engagement: Self-reported

Our results support H3, where we expected that visualisations with SI would be perceived as more engaging than the plain version. The ART model reveals a main effect of the visualisation version ($F(1, 56) = 4.49$, $p = .04$). As shown in Fig. 10-A.2, participants gave the SI version a higher engagement score ($M = 41.67$, $SD = 6.47$) than the plain version ($M = 38.02$, $SD = 8.58$). We also found a main effect of the topic ($F(1, 56) = 4.31$, $p = .04$). In particular, participants were more engaged with the Food topic ($M = 40.88$, $SD = 6.95$) than the Space topic ($M = 38.55$, $SD = 8.56$). The test also revealed an interaction effect ($F(1, 56) = 5.04$, $p = .03$). Further examination revealed that under the SI condition, there is a notable difference in total engagement score between the topics. In particular, participants rated the Space visualisation ($M = 42.04$) to be more engaging than the Food visualisation ($M = 41.30$). On the other hand, under the plain condition, this pattern was reversed; participants perceived the Food visualisation ($M = 40.52$) to be more engaging than the Space visualisation ($M = 35.52$).

Next, we analysed the four dimensions of the UES survey. We found that the visualisation version played an essential factor in aesthetic appeal, similar to Study 1 findings (see Section 5.2.2). Topic, on the other hand, impacted perceived usability. Next, we report these results in detail.

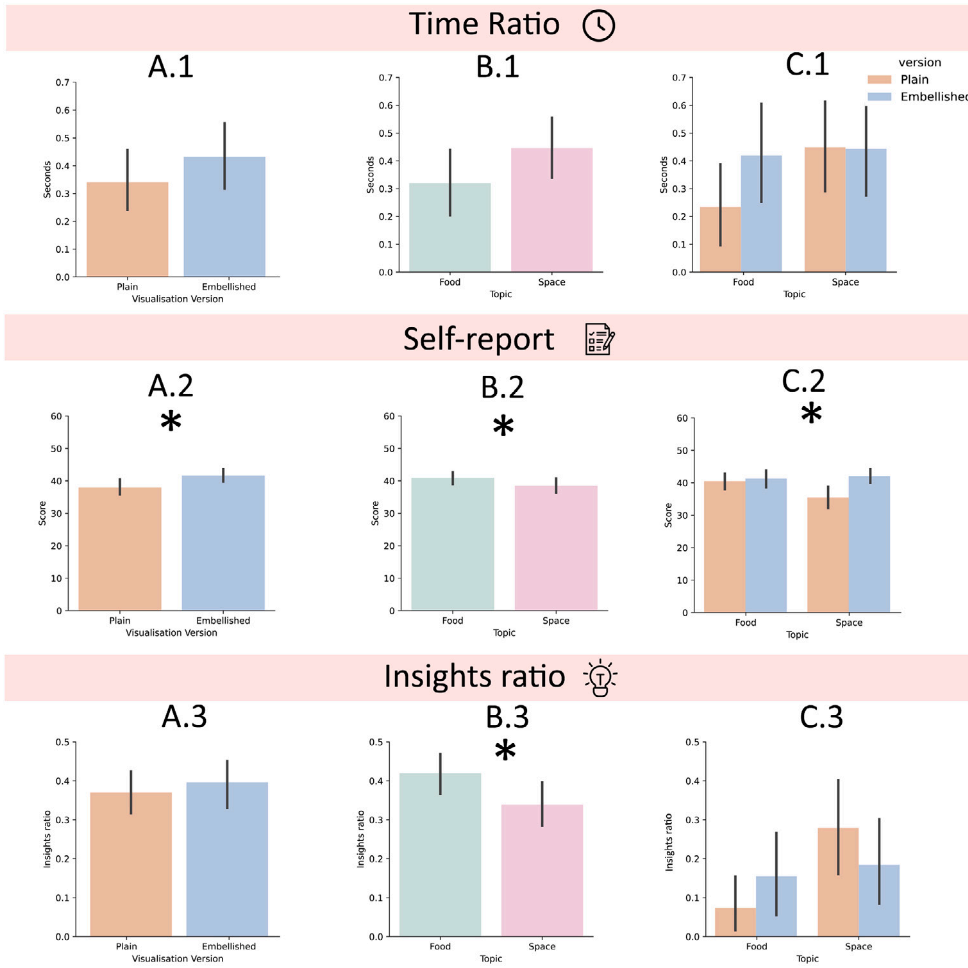


Fig. 10. Results of engagement measures: time ratio (A.1, B.1, C.1), total engagement score (A.2, B.2, C.2), and insights ratio (A.3, B.3, C.3). We organised the results by effects, left-to-right: visualisation version, topic, and interaction effect.

First, for the **aesthetic appeal** dimension, the ART model suggests that there was a main effect of the visualisation version ($F(1,56) = 11.29$, $p = .001$). Participants rated visualisations with SI ($M = 11.02$) as more aesthetically pleasing than plain visualisations ($M = 9.15$). But the topic had no main effect ($F(1,56) = 0.09$, $p = .77$). The test revealed an interaction effect between the visualisation version and the topic ($F(1,56) = 14.99$, $p < .001$). Further analysis suggests that under the SI, the engagement score was notably higher for the Space topic ($M = 11.70$) compared to the Food topic ($M = 10.33$). However, this pattern was reversed under the plain condition where participants rated the Food visualisation ($M = 9.94$) more engaging than the Space visualisation ($M = 8.35$). Second, regarding the **focused attention** dimension, there was no main effect of the visualisation version ($F(1,56) = 0.004$, $p = .95$), no main effect of the topic ($F(1,56) = 0.62$, $p = .43$), and no interaction effect ($F(1,56) = 0.0001$, $p = 0.99$).

Third, for the **perceived usability** dimension, the test reveals no main effect of the visualisation version ($F(1,56) = 2.89$, $p = .09$). However, there was a main effect of the topic ($F(1,56) = 18.46$, $p < .001$). In particular, participants rated the Food visualisation ($M = 11.60$) higher than the Space visualisation ($M = 9.95$) under the perceived usability dimension (evaluates whether reading the visualisation was taxing, frustrating, and confusing). We also found an interaction effect between the visualisation version and the topic ($F(1,56) = 10$, $p = .003$). Further investigation of the interaction revealed that under the SI condition, participants gave a higher engagement score for the Food visualisation ($M = 11.59$) compared to the Space visualisation ($M = 11.11$). Similarly, under the plain condition, participants gave the Food visualisation a higher engagement score ($M = 11.61$) than the Space visualisation ($M = 8.94$). Fourth, regarding the **reward factor** dimension, the test reveals no main effect of the visualisation version ($F(1,56) = 1.46$, $p = .23$), no main effect of the topic ($F(1,56) = 3.82$, $p = .06$), and no interaction effect ($F(1,56) = 0.30$, $p = .58$).

6.2.3. *Cognitive effort: Extracting insights*

The data of one of the participants was excluded from the insights analysis as they did not provide any insights for one of the topics. We report the analysis of the absolute number of insights in Appendix E, which have the same results as the insights ratio.

Insights-ratio. We did not find evidence to support H2.a, which expects the insights ratio (insights from target visualisation/total insights) to be higher for visualisations with SI compared to the plain version. The ART model suggests no main effect of the visualisation version ($F(1,55) = 1.25$, $p = .27$). However, there was a main effect of the topic on the insights ratio ($F(1,55) = 8.28$, $p = .01$) as shown in Fig. 10. The insights ratio was higher for the Food visualisation ($M = 0.43$, $SD = 0.19$) compared to the Space visualisation ($M = 0.34$, $SD = 0.22$). We also found no evidence of an interaction effect ($F(1,55) = 0.43$, $p = .51$).

Effort-level. We did not find evidence to support H2.b, which expects participants to write more low-level and high-level insights for the SI version compared to the plain version. The test suggests that under **low-level insights**, there was no main effect of the visualisation version ($F(1,55) = 1.45$, $p = .23$), no main effect of the topic ($F(1,55) = 0.21$,

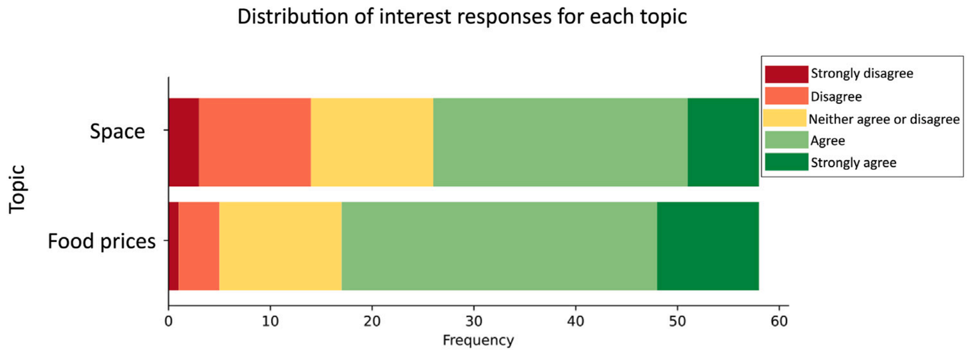


Fig. 11. Distribution of interest responses based on participants' agreement to the following statement: "I am interested in the topic of ⟨*topic*⟩".

Table 3
Quantitative analysis overview of Study 2 based on study factors and DV.

Factor	Dependent variables			
	Visual attention (time-ratio)	Cognitive effort (insights-ratio)	Perceived engagement (self-reported)	Interest
Version (SI, Plain)			[a]	
Topic		[a]	[a]	[a]
Interaction (Topic X Version)			[a]	[a]

[a] Significant difference is based on $\alpha = .05$.

$p = .65$), and no interaction effect ($F(1,55) = 0.75$, $p = .39$). Similarly, under the **high-level insights**, the model suggests no main effect of the visualisation version ($F(1,55) = 0.52$, $p = .48$), no main effect of the topic ($F(1,55) = 1.54$, $p = .22$), and no interaction effect ($F(1,55) = 0.69$, $p = .41$).

6.2.4. *Topic interest*

Fig. 11 presents the distribution of participants' interests across the topics (Food and Space). The ART model suggests no main effect of the visualisation version ($F(1,56) = 2.35$, $p = .13$) on interest. However, the topic had a main effect ($F(1,56) = 6.17$, $p = .02$). Participants were more interested in the topic of food prices than the topic of space. The test also revealed an interaction effect between the topic and the visualisation version ($F(1,56) = 10.64$, $p = .002$). Our results suggest that under the SI condition, interest in the topic was higher for the Space visualisation compared to the Food visualisation. Conversely, under the plain condition, interest in the topic was higher for the Food visualisation compared to the Space visualisation.

We summarise the results and the significant differences for our DV in Table 3. Two important takeaways: (1) The **visualisation version** had a noticeable impact on perceived engagement, particularly on the aesthetic appeal. (2) the **topic** influenced perceived engagement, effort, and interest.

7. Discussion

Our exploratory and two comparative online user studies evaluated the impact of semantic icons on engagement with visualisations. We assessed these visualisations with distracting elements and informational tasks (extracting insights). Next, we discuss the results and how they can be applied in practice, along with future work recommendations.

7.1. *Value of SI in everyday visualisations*

The results of Studies 1 and 2, as summarised in Table 4, indicate that the visualisations with SI were consistently perceived as more engaging than plain visualisations. These results were found when visualisations contained SI by themselves (See Section 5) or accompanied by other elements, such as backgrounds and arrows instead of bars (see Section 6). While our findings in Studies 1 and 2 also demonstrate that the visualisation topic contributes to the perceived engagement (confirming prior work Peck et al., 2019; Kennedy et al., 2016), visualisation designers may have limited control over people's interest in topics. However, they may have more control over the design aspect; therefore, SI may empower them to create engaging visualisations.

Our conclusions about SI confirm our prior findings in Alebri et al. (2023): SI were reported to entice news readers to look at visualisations. Similarly, they confirm Andry et al. (2021) findings where participants perceived the embellished visualisations (various types including SI) more engaging than their plain version. A recurring theme of the studies mentioned above is that participants were asked to assess the level of enticement and engagement of a visualisation. However, in our work, we ask participants to evaluate their level of engagement with a visualisation after extracting insights. Prior work (Graf and Landwehr, 2017) suggests a disparity in evaluation responses based on processing dynamics. In particular, they suggest that assessments of a stimulus after an informational task (e.g., suggesting a title) are perceiver-driven and may yield different results from stimulus-driven assessments (e.g., assess aesthetic appeal) that depend on the participant's immediate reaction. Therefore, our work demonstrates that the positive perception of SI (self-reported) is not limited to the immediate reaction but also persists after an informational task (extracting insights), that requires time and reflection.

Kim et al. (2023) found that integrating human-recognisable objects and graphics in visualisation thumbnails engages and grabs the reader's attention, which could lead to further exploration. The findings of Study 1 and Study 2 extend Kim et al.'s work by demonstrating that visualisations with SI engage viewers with full-size visualisations. Furthermore, the positive findings about the inclusion of SI enhancing perceived engagement challenge (Tufte's 1983) assertion that pictorials are mere 'Chart Junk' with no added value to visualisation.

Further analysis of the perceived engagement (self-reported) results in Studies 1 and 2 suggests that using a simple form of embellishments such as SI improve the aesthetic appeal. Participants in these studies perceived visualisations with SI to be more aesthetically pleasing than plain visualisations (See Sections 5.2.2 and 6.2.2). The seminal work by Bateman et al. (2010) demonstrated that heavily embellished visualisations (found in the wild) that use illustration and visual metaphors

Table 4
An overview of all of the studies (Exploratory study, Study 1, Study 2) results.

Study	Factor	Dependent variables		
		Visual attention (time-ratio)	Cognitive effort (insights-ratio)	Perceived engagement (self-reported)
Exploratory study	Version (SI, P)	*[a]	*[b]	
Study 1	Version (SI, P)			*
	Topic		*	*
	Interaction (Topic X Version)			
Study 2	Version (SI, P)			*
	Topic		*	*
	Interaction (Topic X Version)			*

* Significant difference is based on $p < .05$.
[a] Refer to finding a significant difference for visualisation A_{Gender}.
[b] Refer to finding a significant difference for visualisation B_{Gender}.

were perceived as more attractive than standard visualisations. Still, in that study, the embellished visualisations were created by a renowned designer in a labour-intensive way. The results in our work extend Bateman et al.'s findings by showing that even SI, which can require less labour to produce, could make visualisations to be perceived as more aesthetically pleasing compared to plain visualisations. This is not to imply that the aesthetic appeal level of an extreme embellishment will be similar to SI but to suggest that it is possible to detect a difference with a subtle type of embellishment. Moere and Purchase (2011) assert that visualisation design is not limited to those with extraordinary skills in design, and designers can be taught to make conscious decisions with justifiable reasoning. Our results provide implications to designers aiming to create engaging visualisations with limited resources where they could use icon generation tools (e.g. flaticon,[8] icons8[9]) with their visualisations.

Another value SI bring is enhancing visualisations with less interesting topics. In Study 2, the Space topic was rated to be less interesting than the Food prices topic (refer to Section 6.2.4), which is expected as it is less relevant to most people. In parallel, participants perceived the Space visualisation as less engaging than the Food visualisation (refer to Table 4). However, the results show an interaction effect between the topic and the visualisation version on perceived engagement, which suggests that the Space visualisation became more interesting and engaging with SI. This finding confirms another finding from our prior work (Alebri et al., 2023): participants perceived the SI versions of the Space visualisation enticing, although they felt the topic was irrelevant to them. Our findings extend that work by showing the benefits of SI to less interesting topics even after an informational task (extracting insights), which requires putting some effort and reflecting on the visualisation. This has implications that can benefit visualisation designers keen on expanding their audiences. For instance, prior work (Peck et al., 2019) suggests that people gravitate towards topics that resonate with their personal experiences or are culturally familiar, topics that are distant to them may appear irrelevant to them. Hence, designers may utilise SI to elevate their attention to such topics.

While reliable evidence of the visualisation version's impact on engagement is limited to perceived engagement (self-reported) (see Table 4), SI may still provide a valuable addition to the user experience. For instance, participants consistently in Study 1 and Study 2 perceived the SI version to make the visualisation overall more engaging and more aesthetically pleasing. We also observed that participants in Study 1 perceived the SI version to be more rewarding (see Section 5.2.2), reflecting their interest in the experience and their perception that this version was worth exploring. Sprague and Tory (2012) demonstrated that the casual use of visualisation is not always tied to information acquisition as people could consume visualisation for boredom and entertainment. Wang et al. (2019a) emphasised that the value of visualisation should not be limited to acquiring more knowledge as, for instance, tangible interfaces may not always lead to that but will create a pleasurable experience. Kennedy et al. (2016) suggested multiple ways in which the effectiveness of visualisation in engaging the general public may be defined, such as gaining new confidence in the data and provoking surprise. Similarly, there has been a growing interest in concepts such as enjoyment (Wang et al., 2019b), emotions (Garretón et al., 2023) and empathy (Boy et al., 2017; Dhawka et al., 2023) within the visualisation community. Further investigation is needed to explore how perceived engagement translates to other implicit metrics. Future works could examine whether it leads to sharing, initiating conversations with others, and user loyalty.

7.2. Reflections on engagement measurements

Reflecting on the visual attention results of our primary studies (Study 1 and 2) and the exploratory study (see Table 4), we suggest that employing time ratio (using mouse data) to reveal the distribution of visual attention when distracting elements exist in a view may not work. It should be noted that no effect could be measured in most of the exploratory study visualisations, Study 1, and Study 2 (see Table 4), suggesting that the issue applies equally to more 'subtle' embellishments (Study 1) (refer to Fig. 1) and more 'salient' ones (Study 2) (refer to Fig. 9). Although time spent on interface is a reliable measure of engagement in HCI (Doherty and Doherty, 2018) and employed in visualisation studies (e.g., Boy et al., 2015; Saket et al., 2016; Hung and Parsons, 2017; Wang et al., 2019b), it might be tricky to use in comparative studies evaluating visualisations with SI.

We suspect that both *noise in mouse data* and *the task* we explored influenced our results. *Mouse data* may be noisy because of the sequential inference where mouse data are used to infer gaze data, which are then used to indicate attention (Navalpakkam and Churchill, 2012). Although eye-tracking may seem the next logical step to explore our hypotheses, it may not suit the type of visualisation this work explores and reduce the ecological validity. Eye-tracking may make participants more conscious and slow the viewing process, which may not reflect how the visualisations for the public are viewed. The motivation behind using mouse data was to create a less intrusive experience for the participant where they explore the visualisation on their environment with their regular distractions and desire to procrastinate. Regarding the *task type*, we tracked time spent on a visualisation area while participants extracted insights. This may have motivated them to equally distribute their attention between the visualisations. A type of task that is not information-oriented may produce a different result. For instance, prior work (Haroz et al., 2015) that used a free exploration task

[8] https://www.flaticon.com/.
[9] https://icons8.com/icons.

detected a difference in initial attention between bar charts, stacked pictograph charts, and text. Future work could explore the impact of SI on visual attention while utilising a free exploration task. Furthermore, during the tasks of our primary studies, participants saw two visualisations, another layout may influence the results. For instance, future exploration may adapt (Greussing et al.'s 2020) approach, which not only looked at the time spent on a visualisation area but also the time spent on the subsequent part of the article to understand how the visualisation influenced their attention to the rest of the article. Therefore, we suggest the following RQ: *How do visualisations with SI influence a viewer's attention to subsequent parts of an article?*

Our results also suggest that the alignment of engagement measures (i.e., time ratio, insights ratio, self-report) might be challenging because of the complexity of engagement and its dimensions. The exploratory study and Studies 1 and 2 showed no cases where two measures were aligned because of the visualisation version (see Table 4). For instance, in Studies 1 and 2, the SI version were perceived to be more engaging. However, that did not translate into spending more time on the visualisation version or writing more insights. Wang et al. (2019b) found a similar pattern to the misalignment of engagement measures. They compared engagement of three visualisation styles (comics, infographics, illustrated text) using the UES survey (O'Brien et al., 2018) and found that participants perceived the comic style visualisation as the most engaging. However, when they ran an in-the-wild study by displaying the different styles of visualisations at a conference and tracking the duration visitors spent looking at each style, they did not find a consistent result. Therefore, misalignment of engagement measures means that a specific visualisation design would work for a particular dimension of engagement but not the other and should not dismiss the value the design brings to engaging participants. Using qualitative approaches may further our understanding of the complexity of measuring engagement.

7.3. Limitations

As with every study, some limitations should be considered. First, the perceived engagement (self-reported) results (refer to Table 4) may be limited to simple visualisations. Bar charts, which are common and highly familiar chart types, were used. The number of data points displayed in these visualisations (Study 1 and Study 2) was limited to a small number to ensure that participants would not be overwhelmed by the data. Second, web pages had the distracting item as a data visualisation; other distracting elements may give different results. For instance, de Haan et al. (2017) demonstrated that within a news article, visualisations are not ignored; however, viewers spent significantly more time looking at the article text than the visualisation. Third, participants were asked to extract insights to motivate them to explore the visualisations from their perspective. Other task types could generate different results. Also, no time limit was imposed for the task duration to mimic how they would engage with visualisations in natural settings. Restricting the time participants have to explore the visualisations may influence the results. For instance, Haroz et al. (2015) investigated initial engagement with stacked pictographs by restricting interaction time to two minutes and found that stacked pictographs are more attractive than bars and text.

8. Conclusion

In this work, we reported an exploratory study followed by two comparative online user studies to investigate the impact of visualisations with SI on engagement. Our study design employed distractions to reflect the saliency of embellished visualisation and attention distribution. Participants were asked to extract insights and assess their level of engagement afterwards. Our findings highlight SI's value to the perceived overall engagement, aesthetic appeal, and less interesting and engaging topics. These findings suggest that SI are valuable after performing an informational task, such as extracting insights and reflecting on the visualisation, besides being valuable for the first impression, as prior work indicated. In addition, our visual attention and cognitive effort metrics did not show a consistent pattern of an effect, making it impossible to conclude. Our findings have implications for visualisation designers and storytellers wishing to design engaging visualisations with limited time. We also reflect on our approach and the measurements used to evaluate the visualisations. We hope our work will inspire future exploration of other embellishment types that consider distractions and other aspects that reflect the complexity of the visual environment.

CRediT authorship contribution statement

Muna Alebri: Writing – review & editing, Writing – original draft, Visualisation, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Enrico Costanza:** Writing – review & editing, Validation, Supervision, Software, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Georgia Panagiotidou:** Writing – review & editing, Validation, Supervision, Methodology, Formal analysis, Conceptualization. **Duncan P. Brumby:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization. **Fatima Althani:** Writing – review & editing, Visualization, Resources, Investigation, Conceptualization. **Riccardo Bovo:** Writing – review & editing, Validation, Methodology, Formal analysis, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All of the supplementary material has been cited within the manuscript.

Acknowledgements

This work was supported by the UAE University and the Engineering and Physical Sciences Research Council BIP project (EP/V042327/1).

Appendix A. Exploratory study: Number of insights results

For the Gender view, we found a significant difference between SI and P for **visualisation A** ($U = 196$, $p = .04$, CI95%[0.019, 0.675]) of a small effect ($Cohen's\ d = 0.4$), suggesting that when the visualisation had SI, participants were more inclined (1.68 times) to extract insights compared to the plain version. We did not find evidence of a difference between the SI and the plain version of **visualisation B** ($U = 131$, $p = .52$), **visualisation C** ($U = 200$, $p = .36$), and **visualisation D** ($U = 199$, $p = .99$).

For the Success view, we did not find a difference in the number of insights extracted between SI and P for **visualisation A** ($U = 158$, $p = .78$), **visualisation B** ($U = 119$, $p = .28$), **visualisation C** ($U = 150$, $p = .53$), and **visualisation D** ($U = 182$, $p = .83$).

Appendix B. Study 1: Visual attention results

The Aligned Rank Transform (ART) method was used as the Shapiro –Wilk test of residuals showed that the data deviates from normality ($W = 0.88$, $p < .0001$). The test revealed no main effect of the topic ($F(1,56) = 0.20$, $p = .66$), no main effect of visualisation version ($F(1,56) = 0.003$, $p = .95$), and no interaction effect ($F(1,56) = 1.24$, $p = .27$) on the time ratio (total time spent on a visualisation area divided by time spent on all visualisations).

Appendix C. Study 1: Number of insights results

Regarding the **Number of insights**, the ART model has been used as the Shapiro–Wilk test of residuals showed that the data deviates from normality ($W = 0.97$, $p = .01$). The ART model suggests no main effect of the visualisation version ($F(1,56) = 0.01$, $p = .91$). However, the topic had a main effect on the number of insights extracted from the target visualisations ($F(1,56) = 9.01$, $p = .004$). In particular, participants wrote more insights from the target visualisation under the Sustainability topic ($M = 1.36$, $SD = 1.21$) than the target visualisation in the Gender topic ($M = 0.98$, $SD = 0.96$). The test revealed no interaction between the visualisation version and the topic ($F(1,56) = 1.24$, $p = .27$).

Appendix D. Study 2: Visual attention results

The Aligned Rank Transform (ART) method was used as the Shapiro –Wilk test of residuals showed that the data deviates from normality ($W = 0.86$, $p < .001$). The test revealed no main effect of the topic ($F(1,56) = 1.04$, $p = .31$), no main effect of visualisation version ($F(1,56) = 0.01$, $p = .93$), and no interaction effect ($F(1,56) = 0.55$, $p = .46$) on the time ratio (total time spent on a visualisation area divided by time spent on all visualisations).

Appendix E. Study 2: Number of insights results

First, for the **Number of insights**, the ART model suggests no main effect of the visualisation version ($F(1,55) = 0.59$, $p = .45$), no main effect of the topic ($F(1,55) = 1.88$, $p = .18$), and no interaction effect ($F(1,55) = 0.01$, $p = .91$) on the number of insights extracted from the target visualisation.

References

Adobe, 2021. Color accessability. URL https://color.adobe.com/create/color-accessibility.

Ajani, K., Lee, E., Xiong, C., Nussbaumer Knaflic, C., Kemper, W., Franconeri, S., 2021. Declutter and focus: Empirically evaluating design guidelines for effective data communication. IEEE Trans. Vis. Comput. Graphics 28 (10), 3351–3364. http://dx.doi.org/10.1109/TVCG.2021.3068337.

Akbaba, D., Wilburn, J., Nance, M.T., Meyer, M., 2021. Manifesto for putting 'chartjunk' in the trash 2021!. In: Alt.VIS At IEEE VIS. IEEE, held virtually, pp. 1–5. http://dx.doi.org/10.48550/arXiv.2109.10132, URL http://arxiv.org/abs/2109.10132.

Alebri, M., Costanza, E., Panagiotidou, G., Brumby, D.P., 2023. Embellishments revisited: Perceptions of embellished visualisations through the viewer's lens. IEEE Trans. Vis. Comput. Graphics 30 (1), 1424–1434. http://dx.doi.org/10.1109/TVCG.2023.3326914.

Alebri, M., Costanza, E., Panagiotidou, G., Brumby, D.P., Althani, F., Bovo, R., dataset, 2024. Dataset. URL https://doi.org/10.6084/m9.figshare.25341841.v1.

Amini, F., Henry Riche, N., Lee, B., Leboe-McGowan, J., Irani, P., 2018. Hooked on data videos: Assessing the effect of animation and pictographs on viewer engagement. In: Proceedings of the 2018 International Conference on Advanced Visual Interfaces. Association for Computing Machinery, Castiglione della Pescaia, Italy, pp. 1–9. http://dx.doi.org/10.1145/3206505.3206552.

Anderson, H., Daniels, M., 2016. Film dalogue from 2,000 screenplays, broken down by gender and age. URL https://pudding.cool/2017/03/film-dialogue/.

Andry, T., Hurter, C., Lambotte, F., Fastrez, P., Telea, A., 2021. Interpreting the effect of embellishment on chart visualizations. In: Proc CHI. ACM, Yokohama, Japan, pp. 1–15. http://dx.doi.org/10.1145/3411764.3445739.

Arunkumar, A., Padilla, L., Bae, G.-Y., Bryan, C., 2023. Image or information? Examining the nature and impact of visualization perceptual classification. IEEE Trans. Vis. Comput. Graphics 30 (1), 1–16. http://dx.doi.org/10.1109/TVCG.2023.3326919, URL http://arxiv.org/abs/2307.10571.

Bateman, S., Mandryk, R.L., Gutwin, C., Genest, A., McDine, D., Brooks, C., 2010. Useful junk? The effects of visual embellishment on comprehension and memorability of charts. In: Proc CHI. ACM, Atlanta, pp. 2573–2582. http://dx.doi.org/10.1145/1753326.1753716.

Borgo, R., Abdul-Rahman, A., Mohamed, F., Grant, P.W., Reppa, I., Floridi, L., Chen, M., 2012. An empirical study on using visual embellishments in visualization. IEEE Trans. Vis. Comput. Graphics 18 (12), 2759–2768. http://dx.doi.org/10.1109/TVCG.2012.197.

Borgo, R., Micallef, L., Bach, B., McGee, F., Lee, B., 2018. Information visualization evaluation using crowdsourcing. Comput. Graph. Forum 37 (3), 573–595. http://dx.doi.org/10.1111/cgf.13444.

Borkin, M.A., Bylinskii, Z., Kim, N.W., Bainbridge, C.M., Yeh, C.S., Borkin, D., Pfister, H., Oliva, A., 2016. Beyond memorability: Visualization recognition and recall. IEEE Trans. Vis. Comput. Graphics 22 (1), 519–528. http://dx.doi.org/10.1109/TVCG.2015.2467732.

Borkin, M.A., Vo, A.A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., Pfister, H., 2013. What makes a visualization memorable? IEEE Trans. Vis. Comput. Graphics 19 (12), 2306–2315. http://dx.doi.org/10.1109/TVCG.2013.234.

Boy, J., Detienne, F., Fekete, J.-D., 2015. Storytelling in information visualizations: Does it engage users to explore data? In: Proceedings CHI. ACM, Crossings, Seoul, Korea, pp. 1449–1458. http://dx.doi.org/10.1145/2702123.2702452.

Boy, J., Pandey, A.V., Emerson, J., Satterthwaite, M., Nov, O., Bertini, E., 2017. Showing people behind data: Does anthropomorphizing visualizations elicit more empathy for human rights data? In: Proceedings CHI. ACM, Denver, pp. 5462–5474. http://dx.doi.org/10.1145/3025453.3025512.

Burns, A., Xiong, C., Franconeri, S., Cairo, A., Mahyar, N., 2021. Designing with pictographs: Envision topics without sacrificing understanding. IEEE Trans. Vis. Comput. Graphics 28 (12), 4515–4530. http://dx.doi.org/10.1109/TVCG.2021.3092680.

Chen, M.C., Anderson, J.R., Sohn, M.H., 2001. What can a mouse cursor tell us more? Correlation of eye/mouse movements on web browsing. In: Proc CHI-Short Talks. ACM, pp. 281–282. http://dx.doi.org/10.1145/634067.634234.

Chen, Y., Yang, J., Ribarsky, W., 2009. Toward effective insight management in visual analytics systems. In: PacificVis Proceedings. IEEE, Beijing, China, pp. 49–56.

Choi, J., Oh, C., Kim, Y., Kim, N.W., 2023. VisLab: Enabling visualization designers to gather empirically informed design feedback. In: Proceedings CHI. ACM, Hamburg, Germany, pp. 1–18. http://dx.doi.org/10.1145/3544548.3581132.

Cleveland, W.S., McGill, R., 1984. Graphical perception: Theory, experimentation, and application to the development of graphical methods. Am. Stat. Assoc. 79 (387), 531–554. http://dx.doi.org/10.2307/2288400.

de Haan, Y., Kruikemeier, S., Lecheler, S., Smit, G., van der Nat, R., 2017. When does an infographic say more than a thousand words? J. Stud. 19 (9), 1293–1312. http://dx.doi.org/10.1080/1461670X.2016.1267592.

Dhawka, P., He, H.A., Willett, W., 2023. We are the data: Challenges and opportunities for creating demographically diverse anthropographics. In: Proceedings CHI. ACM, Hamburg, Germany, pp. 1–14. http://dx.doi.org/10.1145/3544548.3581086.

Doherty, K., Doherty, G., 2018. Engagement in HCI: Conception, theory and measurement. ACM Comput. Surv. 51 (5), 1–39. http://dx.doi.org/10.1145/3234149.

Garretón, M., Morini, F., Celhay, P., Dörk, M., Parra, D., 2023. Attitudinal effects of data visualizations and illustrations in data stories. IEEE Trans. Vis. Comput. Graphics 1–16. http://dx.doi.org/10.1109/TVCG.2023.3248319.

Government, U., 2022. The-national-minimum-wage-in-2022 @ www.gov.uk. URL https://shorturl.at/fhENR.

Graf, L.K., Landwehr, J.R., 2017. Aesthetic pleasure versus aesthetic interest: The two routes to aesthetic liking. Front. Psychol. 8 (15), 1–15. http://dx.doi.org/10.3389/fpsyg.2017.00015.

Greussing, E., Kessler, S.H., Boomgaarden, H.G., 2020. Learning from science news via interactive and animated data visualizations: An investigation combining eye tracking, online survey, and cued retrospective reporting. Sci. Commun. 42 (6), 803–828. http://dx.doi.org/10.1177/1075547020962100.

Haroz, S., Kosara, R., Franconeri, S.L., 2015. ISOTYPE visualization – working memory, performance, and engagement with pictographs. In: Proceedings CHI. ACM, Seoul, pp. 1191–1200. http://dx.doi.org/10.1145/2702123.2702275.

He, H.A., Walny, J., Thoma, S., Carpendale, S., Willett, W., 2023. Enthusiastic and grounded, avoidant and cautious: Understanding public receptivity to data and visualizations. IEEE Trans. Vis. Comput. Graphics 1–11. http://dx.doi.org/10.1109/TVCG.2023.3326917.

Hickey, W., 2014. The dollar-and-cents case against hollywood's exclusion of women. URL https://53eig.ht/31dCKZc.

Huang, J., White, R.W., Dumais, S., 2011. No clicks, no problem: Using cursor movements to understand and improve search. In: Proceedings CHI. ACM, Vancouver, Canada, pp. 1225–1234. http://dx.doi.org/10.1145/1978942.1979125.

Hullman, J., Adar, E., Shah, P., 2011. Benefitting InfoVis with visual difficulties. IEEE Trans. Vis. Comput. Graphics 17 (12), 2213–2222. http://dx.doi.org/10.1109/TVCG.2011.175.

Hung, Y.H., Parsons, P., 2017. Assessing user engagement in information visualization. In: Proceedings CHI Extended Abstracts. ACM, Denver, USA, pp. 1708–1717. http://dx.doi.org/10.1145/3027063.3053113.

IMDb.com, 2020. IMDb datasets. URL www.imdb.com/interfaces.

International, S.T., 2018. Carbon footprint of tourism. URL https://rb.gy/oe328w.

Jamieson, S., 2004. Likert scales: How to (ab)use them. Med. Educ. 38, 1217–1218. http://dx.doi.org/10.1111/j.1365-2929.2004.02012.x.

Kennedy, H., Hill, R.L., Allen, W., Kirk, A., 2016. Engaging with (big) data visualizations: Factors that affect engagement and resulting new definitions of effectiveness. First Monday 21 (11), 1–21. http://dx.doi.org/10.5210/fm.v21i11.6389, URL https://shorturl.at/efFKS.

Kim, H., Kim, J., Han, Y., Hong, H., Kwon, O.S., Park, Y.W., Elmqvist, N., Ko, S., Kwon, B.C., 2023. Towards visualization thumbnail designs that entice reading data-driven articles. IEEE Trans. Vis. Comput. Graphics 30 (1), 1–16. http://dx.doi.org/10.1109/TVCG.2023.3278304.

Lagun, D., Lalmas, M., 2016. Understanding and measuring user engagement and attention in online news reading. In: Proceedings of the Ninth ACM International Conference on Web Search and Data Mining. Association for Computing Machinery, New York, NY, USA, pp. 113–122. http://dx.doi.org/10.1145/2835776.2835833.

2024. Launch vehicle. http://dx.doi.org/10.2307/j.ctvqsdzvs.9, URL https://en.wikipedia.org/wiki/Launch_vehicle#.

Lee, B., Choe, E.K., Isenberg, P., Marriott, K., Stasko, J., 2020. Reaching broader audiences with data visualization. IEEE Comput. Graph. Appl. 40 (2), 82–90. http://dx.doi.org/10.1109/MCG.2020.2968244.

Liu, S., 2023. Individual actions vacationers say they take for sustainability. URL https://shorturl.at/cvQY7.

Mahyar, N., Kim, S.-H., Kwon, B.C., 2015. Towards a taxonomy for evaluating user engagement in information visualization. In: Personal Visualization : Exploring Data in Everyday Life At IEEE VIS. IEEE, Chicago, USA, pp. 1–4.

Moere, A.V., Purchase, H., 2011. On the role of design in information visualization. Inf. Vis. 10 (4), 356–371. http://dx.doi.org/10.1177/1473871611415996.

N, E., 2023. Data analysis of female presentation in movies — Has the # of bechdel test-passing movies increased?—. URL https://shorturl.at/uHOUY.

Navalpakkam, V., Churchill, E.F., 2012. Mouse tracking: Measuring and predicting users' experience of web-based content. In: Proceedings CHI. ACM, Texas, USA, pp. 2963–2972. http://dx.doi.org/10.1145/2207676.2208705, URLA.

O'Brien, H.L., Cairns, P., Hall, M., 2018. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. Int. J. Human Comput. Stud. 112, 28–39. http://dx.doi.org/10.1016/j.ijhcs.2018.01.004.

O'Brien, H.L., Toms, E.G., 2010. The development and evaluation of a survey to measure user engagement. J. Am. Sco. Inf. Sci. Technol. 61 (1), 50–69. http://dx.doi.org/10.1002/asi.

Pandey, A.V., Rall, K., Satterthwaite, M.L., Nov, O., Bertini, E., 2015. How deceptive are deceptive visualizations?: An empirical analysis of common distortion techniques. In: Proceedings CHI. ACM, Seoul, Korea, pp. 1469–1478. http://dx.doi.org/10.1145/2702123.2702608.

Peck, E.M., Ayuso, S.E., El-Etr, O., 2019. Data is personal: Attitudes and perceptions of data visualization in rural Pennsylvania. In: Proceedings CHI. ACM, Glasgow, pp. 1–12. http://dx.doi.org/10.1145/3290605.3300474.

Peña, A., Ragan, E.D., Harrison, L., 2020. Memorability of enhanced informational graphics: The effects of design relevance and chart type on recall. Interdiscip. J. Signage Wayfinding 4 (1), 19–33. http://dx.doi.org/10.15763/issn.2470-9670.2020.v4.i1.a54.

Pousman, Z., Stasko, J.T., Mateas, M., 2007. Casual information visualization: Depictions of data in everyday life. IEEE Trans. Vis. Comput. Graphics 13 (6), 1145–1152. http://dx.doi.org/10.1109/TVCG.2007.70541.

Race, M., 2023. Why is inflation higher in UK than other countries? URL https://www.bbc.co.uk/news/business-65037292.

Saket, B., Endert, A., Demiralp, C., 2019. Task-based effectiveness of basic visualizations. IEEE Trans. Vis. Comput. Graphics 25 (7), 2505–2512.

Saket, B., Scheidegger, C., Kobourov, S., 2016. Comparing node-link and node-link-group visualizations from an enjoyment perspective. In: Computer Graphics Forum (EuroVis). Eurographics Association, pp. 41–50. http://dx.doi.org/10.1111/cgf.12880.

Setlur, V., Mackinlay, J.D., 2014. Automatic generation of semantic icon encodings for visualizations. In: Proceedings CHI. Toronto, ON, Canada, pp. 541–550. http://dx.doi.org/10.1145/2556288.2557408.

Shi, Y., Liu, P., Chen, S., Sun, M., Cao, N., 2022. Supporting expressive and faithful pictorial visualization design with visual style transfer. IEEE Trans. Vis. Comput. Graphics 29 (1), 236–246. http://dx.doi.org/10.1109/TVCG.2022.3209486.

Simons, D.J., Chabris, C.F., 1999. Gorillas in our midst: sustained inattentional blindness for dynamic events. Perception 28 (9), 1059–1074. http://dx.doi.org/10.1068/p281059.

Skau, D., Kosara, R., 2017. Readability and precision in pictorial bar charts. In: Computer Graphics Forum -Short Paper. Eurographics Association, Barcelona, pp. 91–95. http://dx.doi.org/10.2312/eurovisshort.20171139.

Sprague, D., Tory, M., 2012. Exploring how and why people use visualizations in casual contexts: modeling user goals and regulated motivations. Inf. Vis. 11 (2), 106–123. http://dx.doi.org/10.1177/1473871611433710.

Tufte, E., 1983. The Visual Display of Quantitative Information. Graphics Press, USA, http://dx.doi.org/10.5555/33404.

Vande Moere, A., Tomitsch, M., Wimmer, C., Christoph, B., Grechenig, T., 2012. Evaluating the effect of style in information visualization. IEEE Trans. Vis. Comput. Graphics 18 (12), 2739–2748. http://dx.doi.org/10.1109/TVCG.2012.221.

Wang, Y., Segal, A., Klatzky, R., Keefe, D.F., Isenberg, P., Hurtienne, J., Hornecker, E., Dwyer, T., Barrass, S., 2019a. An emotional response to the value of visualization. IEEE Comput. Graph. Appl. 39 (5), 8–17. http://dx.doi.org/10.1109/MCG.2019.2923483.

Wang, Z., Wang, S., Farinella, M., Murray-Rust, D., Riche, N.H., Bach, B., 2019b. Comparing effectiveness and engagement of data comics and infographics. In: Proceedings CHI. ACM, Glasgow, UK, pp. 1–12. http://dx.doi.org/10.1145/3290605.3300483.

Wood, J., Isenberg, P., Isenberg, T., Dykes, J., Boukhelifa, N., Slingsby, A., 2012. Sketchy rendering for information visualization. IEEE Trans. Vis. Comput. Graphics 18 (12), 2749–2758. http://dx.doi.org/10.1109/TVCG.2012.262.