Bridging Educational Theories of Cognitive Load to Visualization Design and Evaluation

Anne-Flore Cabouat*
Université Paris-Saclay, Inria, LISN, Télécom Paris, Institut Polytechnique de Paris, i3, CNRS

Lorenzo Ciccione†
DysCo Lab, Université Paris 8, Collège de France, Université Paris Sciences Lettres (PSL)

Samuel Huron‡
Télécom Paris, Institut Polytechnique de Paris, i3, CNRS

Tobias Isenberg*
Université Paris-Saclay, Inria, LISN, CNRS

Petra Isenberg*
Université Paris-Saclay, Inria, LISN, CNRS

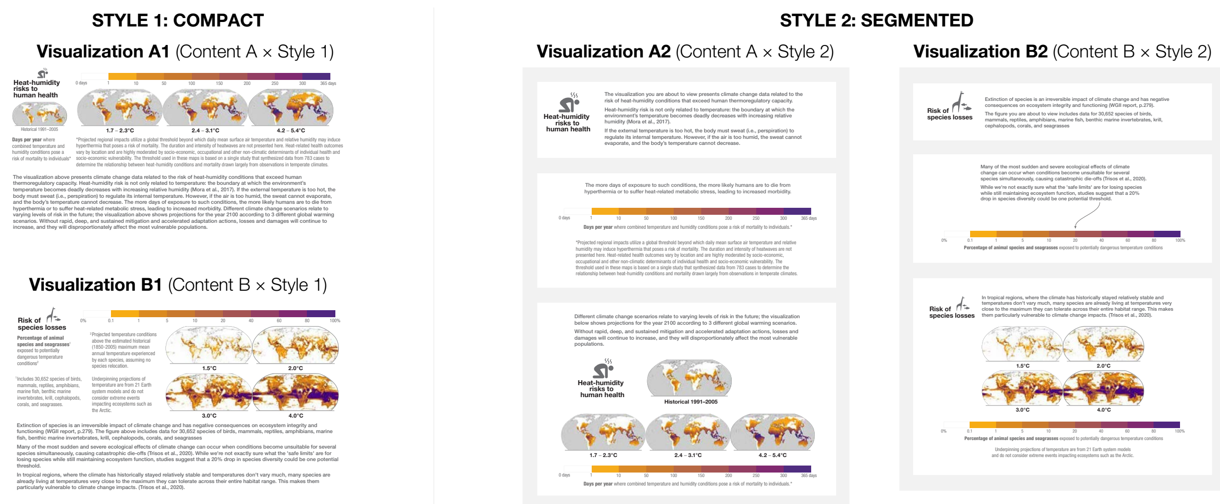


Figure 1: In our study, we presented participants with two visualizations and text extracted (Contents A and B above) from the IPCC Synthesis Report Summary for Policy Makers [19] (specifically from Figure SPM3). We showed the same content using two different presentation styles: compact (Style 1) and segmented (Style 2). Participants saw either A1 and B2, or A2 and B1.

Abstract

We explore the validity and applicability of educational and cognitive science theoretical frameworks for designing and evaluating data visualizations. Specifically, we are interested in using well-known frameworks from other domains to learn about how the subjective readability of a visualization relates to the perceived cognitive load required to acquire knowledge from it. To that end, we conducted an online randomized study in which each participant performed learning tasks on two different data visualizations. One was presented in three successive parts, following the *segmenting* principle from the Cognitive Theory of Multimedia Learning, and the other was presented as a single image. Although most learners preferred the segmented style, this treatment did not significantly affect the overall mental effort they reported. Subjective measures of *extraneous* cognitive load, however, significantly and negatively correlated with visualizations' perceived readability measures. In other words, if a learner found a visualization more readable, they felt it required less mental effort to parse relevant information from it for learning. In addition to a qualitative analysis of learners' preferences, we also contribute an interdisciplinary perspective on cognitive processing of visualizations and a discussion of implications for designing and evaluating data visualizations beyond educational contexts.

Index Terms: Cognitive load, readability, multimedia learning.

1 Introduction

Researchers in Human-Computer Interaction [8], as well as in data visualization specifically [30], have advocated for a deeper integration of insights from external disciplines' theories. How people can gain knowledge from visual data representations has been studied by researchers from various domains such as education [24, 75], cognitive psychology [28, 34], health decision making [31], computer science [11, 15], neuroscience [20, 21], and medias studies [70]. In this work, we build upon previous research that connects educational science to the design and evaluation of visualizations (e.g., [1, 5, 16, 37, 77]) to inform how we conceptualize and assess users' experiences with visualizations from a learning science perspective.

We propose using concepts from the cognitive psychology of learning as conceptual lenses to better describe and support readers' cognitive processes in data visualization. Specifically, as illustrated in Fig. 2, we focus on the Cognitive Load Theory (CLT) [80] and the Cognitive Theory of Multimedia Learning (CTML) [51], and how they might help us to study the readability of data visualizations, which we broadly define as "how easy it is for people to retrieve useful information from a visual representation of data" [17].

CLT is organized around the concept of *cognitive load*, which represents the workload a learning task imposes on the learner's cognitive system [62]. CTML provides us with a theoretical framework for designing instructional messages that can support meaningful

*e-mail: given-name.family_name@inria.fr
†e-mail: given_name.family_name@cri-paris.org
‡e-mail: given_name.family_name@telecom-paris.fr

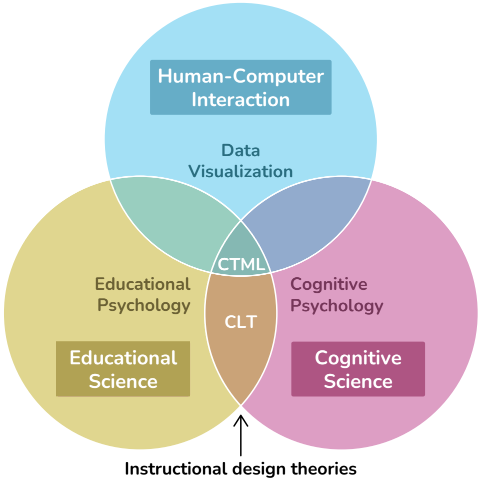


Figure 2: We propose to leverage concepts from the instructional design literature—specifically, from the Cognitive Load Theory (CLT) and the Cognitive Theory of Multimedia Learning (CTML)—as conceptual lenses to study communicative data visualization.

learning—i. e., improve learners' memorization and understanding. In the first part of this work, we describe components from both theories and propose connections to existing concepts in data visualization design and evaluation. To explore how borrowing from CLT and CTML can help us learn about people's experience of visualizations, we then conduct a study. In the study we examine the effect of applying a specific instructional design principle from CTML on both the perceived readability of a data visualization and the subjective cognitive load experienced by readers, as interpreted through the components described in CLT.

2 Related work

We are not the first to be interested in applying theories from educational science to data visualization [74]. There is a growing body of work proposing, for instance, to adapt Bloom's taxonomy of educational goals to identify and assess different levels of understanding in data visualization viewers [16] and to determine visualization task complexity levels [58]. Another stream of research proposes using learning objectives as a specification framework for designing and evaluating visualizations [1, 49]. Research on visualization onboarding systems [77] and constructive visualization [38] is also clearly grounded in educational theories [78]. Beyond these publications that explicitly draw from educational sciences, a few works in our field focus on visualization understanding, recall, and recognition. Such studies examine the memorability of data visualizations [10, 11], effective design strategies for supporting readers' understanding and retention of information from a data visual representation [9], or how people make sense of unfamiliar visualizations [47]. Although the aforementioned studies were not conducted in educational settings, their measured outcomes are comparable to learning goals as defined in the context of multimedia learning [40]: **remembering** (i. e., retention) and **understanding** (i. e., transfer). Educational theories are thus likely to provide us with useful insights into the visualization field at large, which motivated the work we present here.

2.1 Cognitive load and instructional design

Cognitive Load Theory (CLT) explains that a learner must allocate cognitive resources to a learning task in order to acquire knowledge or develop a skill [62]. "Cognitive load," in this context, refers to the total amount of cognitive resources required for a task. CLT also relies on the assumption that a learner's pool of cognitive resources cannot be increased, and that learning cannot occur effectively if the demands of the task exceed the learner's available cognitive resources. There are three main cognitive load components in CLT: intrinsic load, extraneous load, and germane load—also called relevant information load.

Intrinsic load is imposed by the nature and basic structure of the information that a learner needs to process *simultaneously* to achieve a learning goal. It relates to the notion of information complexity, known in CLT as "element interactivity": the number of interacting pieces of information that the learner must process together. While most measures of information complexity refer purely to the characteristics of information, from a CLT perspective, intrinsic load is heavily influenced by the *learner's prior knowledge* on the topic at hand. This is because knowledge acquisition facilitates the integration of multiple pieces of information into single, more manageable units [61]. For example, a kid who is learning to read aloud will have to put effort in processing the individual letters "a", "n", and "y", while a more experienced reader will be able to process and articulate the word "any" as a single entity. In CLT, intrinsic load is completely independent of the form in which the information is presented. As a result, the intrinsic load of a learning task cannot be altered other than by increasing the learner's prior knowledge or changing the learning goal itself.

Extraneous load, in contrast, is solely imposed by the design of learning material and activities. It stems from the mental resources a learner must allocate to processing the representation of information, and is thus influenced by how efficient this representation is to convey relevant information to a specific learner. In addition, processing any piece of information that is NOT relevant to the learning goal also adds to extraneous load.

Germane load—also referred to as **relevant information load** by some authors [61]—represents the mental effort a learner invests in knowledge acquisition itself—i. e., updating their mental representations of the world with new, relevant information. It is thus closely linked to intrinsic load, as more complex information often requires greater cognitive engagement for meaningful learning. Germane load is also thought of as having a redistributing function [82]: rather than adding to overall load, germane load reflects a beneficial use of cognitive resources, ideally redirected from unnecessary processing (i. e., extraneous load) toward activities germane to learning.

Since learners have a limited pool of cognitive resources, learning can only effectively occur if the sum of intrinsic and extraneous cognitive loads remains lower than the learner's cognitive capacity limit, so that the learner can dedicate resources for acquiring new, relevant information. Thus, CLT suggests that efficient *instructional design* should aim to reduce extraneous cognitive load, enabling learners to allocate a greater portion of their cognitive resources to issues that are directly relevant to the learning goal, rather than to processing extraneous information or inefficient representations of relevant information. In this context, instructional design refers to the planning of pedagogical strategies and the design of effective learning materials and activities to support specific learning goals. Fig. 3 presents 4 different learning scenarios along with examples of instructional design strategies to prevent cognitive overload.

The concept of cognitive load is not new to visualization researchers, although it is not often discussed as such. In particular, electroencephalography-based studies have explored how different visualization formats affect cognitive load [3, 89]. While some researchers focused on isolating extraneous cognitive load by controlling for germane load and intrinsic load in their experimental design [3], others did not attempt to obtain distinct measures of extraneous, intrinsic, and germane load [89]. As CLT predicts the effects of instructional design manipulations on specific types of cognitive load, we included distinct measurements for overall, intrinsic, extraneous, and germane cognitive loads in our study design.

2.2 Cognitive Theory of Multimedia Learning

For the past 25 years, the Cognitive Theory of Multimedia Learning [54] (CTML) has been a framework for integrating, refining, and

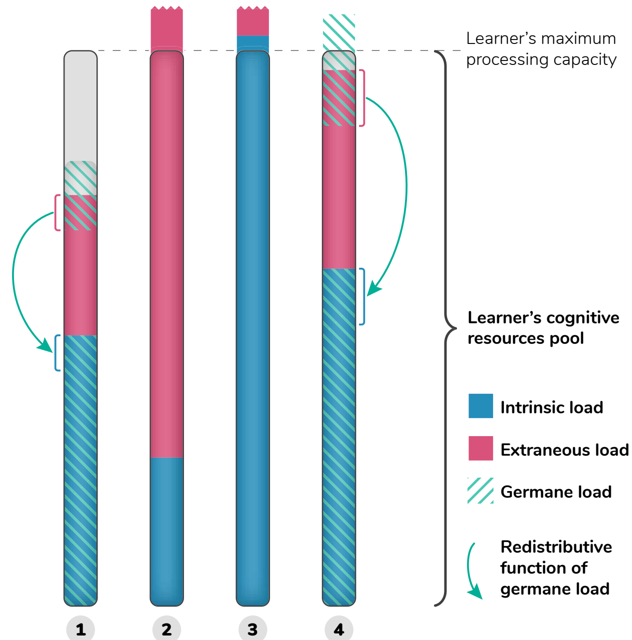


Figure 3: Four learning scenarios in CLT.

Case 1: After accounting for intrinsic and extraneous load, the learner has sufficient cognitive resources available to filter out irrelevant information and engage in effective learning.

Case 2: Extraneous load is too high—the learner is overwhelmed by the way information is presented or by irrelevant details. *Example solutions*: remove irrelevant content, find a representation that is easier to process for this learner, or improve the learner's ability to process the chosen type of representation (e. g., in data visualization contexts, by increasing the learner's idiom-specific visualization literacy, or their ability to leverage the system's interactive features).

Case 3: The knowledge to acquire is intrinsically too complex for this learner, leading to cognitive overload and loss of information. *Example solutions*: build the learner's prior knowledge on the topic or adjust the learning goal.

Case 4: Once intrinsic and extraneous loads have been addressed, the learner is operating near their cognitive capacity. Their available bandwidth for germane processing is constrained, possibly hindering their ability to redirect cognitive resources from extraneous to intrinsic processing. *Example solutions*: if extraneous load has already been reduced to a minimum, provide scaffolds for germane processing, e. g., by breaking the learning task into smaller steps or by supporting meta-cognitive strategies.

extending CLT and other cognitive and learning theories within the specific context of multimedia learning. Multimedia instructional material is broadly defined in CTML as "a communication containing words and pictures intended to foster learning" [53]. Although initially focused on pictorial and diagrammatic representations, researchers working in the CTML paradigm have also investigated data charts and graphs (e. g., [43, 73]). Data visualizations commonly use both graphical and textual content, thus align with CTML's scope.

The extensive empirical work on CTML resulted in an evolving set of design principles and provides an elaborate framework dedicated to multimedia instructional design. The main design principle in CTML is the multimedia principle [52], which echoes Paivio's dual-coding theory [23, 64] in stating that people learn better from words and pictures than from words alone. Other principles in CTML define the conditions to optimize multimedia learning outcomes [54]. Those guidelines aim to reduce extraneous cognitive load, help the learner focus on relevant information processing (i. e., support germane cognitive load), foster engagement and active learning, and integrate social and affective components in instructional design.

Visualization researchers are increasingly applying instructional design principles from CTML such as the *cueing*—or *signaling*—principle [84] to develop effective visualization guidance systems (e. g., [7, 72]). Several relevant principles, however, remain less explored within the visualization community. One such example is the *expertise reversal* principle [42, 41], which emphasizes the importance of adapting multimedia content to the learner's level of expertise. While novice learners often benefit from structured guidance that simplifies problem-solving, the same support may hinder more advanced learners by limiting their autonomy. This principle resonates with visualization research focused on characterizing the needs of novice and expert users (e.g., [88, 47]).

We explore the application of the *segmenting* principle as a strategy for helping data visualization readers manage their cognitive load. By breaking complex information into smaller, user-controllable segments, this principle aims to reduce extraneous load and enable learners to allocate more cognitive resources to relevant information processing at their own, self-regulated pace. Although this effect is well-studied in multimedia learning, its application in data visualization has been limited. Prior work, moreover, has primarily focused on onboarding scenarios—helping users learn how to interpret novel visualizations (e.g., [79, 45])—rather than supporting them as they engage with familiar visualizations under cognitive load.

On the educational literature side, research in multimedia learning often focuses on visual representations, such as diagrams, to support learning [56]. Visual data representations like charts or graphs, however, are discussed much less frequently. Even when studying learning activities with data visualizations, cognitive load measurements are rare. For instance, a study [50] showed that visual cues and added explanatory visual material could increase the number of relational and causal statements that learners can infer from a graph. Yet the authors focused on learning outcomes without examining whether these improvements were linked to reduced cognitive load as predicted by CTML principles. Measuring outcomes alone can show that learning occurred, but without cognitive load evaluation it remains unclear why the design was effective or how it might be adapted for different learners or contexts. To help bridge this gap, we focused on cognitive load evaluation rather than learning outcomes. Understanding *how* and *why* design choices affect knowledge acquisition—through their impact on mental effort—is essential for developing effective, generalizable instructional strategies.

If we think of data visualization as a learning problem [1], we can use CTML principles to improve the design of data visual representations for communicative purposes. We can also evaluate visualization activities as learning experiences using cognitive load measurements. Although visualization researchers [1, 36] proposed somewhat similar ideas, they did not put forward a theoretical analysis of how various characteristics of data visualizations might affect different types of cognitive load in viewers. To bridge this gap, we propose such an analysis and examine how we can connect concepts from the CLT and CTML framework to data visualization literature.

3 Factors of Cognitive Load in Data Visualization

CLT is, in essence, relevant to the presentation of any complex information that requires a reduction in cognitive workload [81]. Early work on visualization complexity showed that visual representations of data can deliver overwhelming amounts of information [13]. From a visualization viewpoint, some of this complexity can be attributed to the nature of the underlying data, some to the appropriateness and efficiency of the visual design, some to the viewer's knowledge on the topic, and some to their data visualizations reading skills [39]. Conversely, in CLT and the CTML the notion of complexity relates only to the core of informational content.

SUMMARY: KEY CONCEPTS FROM COGNITIVE LOAD THEORY

Cognitive load: The total amount of cognitive resources required for a knowledge acquisition task.

Intrinsic load: The load imposed on the learner based on the complexity of the relevant information, regardless of how it is presented. The complexity of relevant information also varies according to each individual learner's prior knowledge.

Extraneous load: The load imposed by the design of learning material and activities, regardless of the intrinsic complexity of information, and by the processing of irrelevant information.

Germane load: The cognitive resources a learner invests in knowledge acquisition itself, and in distributing mental effort between relevant and irrelevant processing.

Instructional design: The planning of pedagogical strategies and the design of effective learning materials and activities to support specific learning goals.

In this context, *element interactivity* is described in CLT as the amount of novel elements of information that a learner needs to process simultaneously to achieve the learning goal. The level of relevant element interactivity is the primary factor that determines **intrinsic load**: for a given learning task with a specific learner's level of prior knowledge on the topic, intrinsic load is fixed [61]. When it comes to data visualizations, underlying datasets are usually too large to be held in a learner's memory—in fact, visual representations have early-on been described as a form of external memory [46]—but a reader's *domain knowledge* could mediate intrinsic load. Regardless of the domain, data visualizations also frequently present information with high element interactivity because they depict relations between multiple variables, i. e., *data dimensions*, and/or across a large number of entities, or *data points*. If all represented data dimensions and data points are relevant to the learning goal, increasing their number would directly lead to higher intrinsic cognitive load in data visualization.

Such dataset elements, however, are not always necessary to achieve a given learning task. As the inclusion of non-essential elements in learning materials can lead to unnecessary cognitive effort for the learner, a visualization showing superfluous data dimensions or points will generate **extraneous load**. Another essential data visualization quality criterion relating to extraneous load is *readability*, which broadly refers to the ease with which a reader can visually retrieve information from a visualization [18, 32, 83]. Visual design choices significantly impact readability. For instance, not all visual channels—such as color, position, size, or shape—used to represent data are equally *expressive* (i. e., effective) [59]. Furthermore, visualization effectiveness is task-dependent [65]. This means that how well a representation supports relevant visualization *tasks* [2] will also affect readability and, consequently, extraneous cognitive load. The characteristics of learners also play a crucial role in readability [90]. In particular, learners with low *visualization literacy* [12, 48] may struggle to interpret data representations, thus experiencing lower readability [17]—and higher extraneous cognitive load. Overall, we can expect a data visualization to generate less extraneous load when it is more readable. Our study includes subjective measures of extraneous load and readability to help bridge the gap between visualization research and CLT and CTML. A strong correlation between participants' ratings of extraneous load and perceived readability would indicate that instructional theories can provide a relevant framework for evaluating data visualizations.

Beyond this list of readability-related characteristics, digital representations of data often feature *interactive* elements, which can also affect the learner's cognitive load. Poorly designed interactive features will increase extraneous cognitive load again. Conversely, well-designed features can provide learners with greater control over the pace of information display and on-demand access to more detailed information [71]—echoing Shneiderman's mantra "Overview first, zoom and filter, and details on-demand" [76]. According to CTML, giving learners control over instructional content level and pacing can help learners to better manage the allocation of cognitive resources to **germane load** [55]. More specifically, multiple studies [67] have shown that *segmenting* the presentation of information in smaller chunks and allowing learners to control the pace of display can aid learning and reduce the overall cognitive load. The segmenting principle finds a parallel in data comics [6]—visualization storytelling formats using sequential panels and data-driven narratives to guide the reader. Constructive data comics [87], in particular, explain the visual encoding choices, supporting readers in building conceptual understanding of a representation. By sequentially unfolding information across panels, data comics implement a form of segmentation that may help reduce extraneous cognitive load. However, data comics also leverage storytelling, visual cueing, and graphical conventions from traditional comics, making it difficult to isolate the effects of segmentation. To examine the specific impact of segmentation in a data visualization context, we compare cognitive load ratings between a segmented visualization and a non-segmented control.

To summarize, we found strong indications in the literature that Cognitive Load Theory (CLT) should provide a relevant framework to assess visualization efficiency. In addition, the related Cognitive Theory of Multimedia Learning (CTML) should also provide relevant design principles for improving data visualizations. We thus want to address the following research question: **To what extent do CLT and CTML apply to information retrieval from data visualizations?** To answer it, we derived two hypotheses from the reviewed literature:

- **H1:** When learning from data visualizations, extraneous cognitive load and data visualization readability are negatively correlated.
- **H2:** Applying the segmenting principle to data visualizations decreases the overall cognitive load in learners.

4 Method

We conducted a pre-registered `osf.io/ptsne` cross-sectional experimental study with a within-subject design. Our goal was to assess the correlation between perceived visualization readability and learners' subjective extraneous cognitive load and to assess the effects of applying the CTML segmenting principle to the presentation of data visualizations on learners' cognitive load. We collected subjective ratings of readability and cognitive load from learners on two different data visualization stimuli: a control stimulus, and a segmented stimulus. To gain a more comprehensive understanding of our participants' experiences with the visualizations, we also collected their feedback about which visualization their preferred and why.

4.1 Participants

To recruit participants, we disseminated a message to our personal and academic international networks via e-mail, Discord, Facebook, and WhatsApp groups. The invitation message contained a short description of the study and stated that participants would not receive compensation. Participants were required to be fluent English speakers, of legal age in their country of residency, to have normal or corrected to normal vision, and to use a computer device to ensure a minimum display width of 1024 pixels. As pre-registered, we closed the survey after 10 days because we had reached our minimum sample size target of 20 participants. 34 participants fully completed the survey, although three were unable to express their visualization preferences in the last question due to a technical issue; two others chose not to explain their preferences. As a result, five participants were not included in the qualitative analysis, but their other answers in the survey were still used for quantitative analyses.

All age groups from 18−24 to 65+ were represented, but most participants belonged to the 25–34 age group ($N = 22$). 18 participants identified as men (cis or transgender), 15 as women (cis or transgender), and one as non-binary. Most of the participants had completed a Master's degree ($N = 21$) or a Bachelor's degree ($N = 9$). The majority resided in Europe ($N = 23$, including 16 people living in France), five participants resided in Asia, three in Africa, two in North America, and one in Oceania. Table 5 in App. D.1 reports all collected demographics. Participants' mean self-reported familiarity with the topic of climate change was 3.76 ($SD = 1.05$) on a 7-point Likert scale (1 = *I don't know what climate change is* to 7 = *I am a climate scientist myself*). Most participants reported they *rarely (a few times a year)* ($N = 6$), *sometimes (1–3 times a month)* ($N = 16$), or *frequently (once a week or more)* ($N = 8$) read or watched content related to climate change; four participants reported having already created content related to the topic (see Tabs. 6 and 7 in App. D.2 for details).

4.2 Materials

Learning tasks. Participants had to complete three different learning tasks for each visualization: (1) write one to five *key takeaways* they could learn from the visualization, (2) write one to five *questions* they could answer using the visualization, and (3) select *correct statements* from a list of 6, in which there were 3 false and 3 correct. We derived these tasks from previous work [66], which indicated that different tasks elicit different response patterns in people, in line with existing survey design recommendations [44].

Data visualizations. Each participant completed learning tasks on two different data visualizations (see teaser Figure 1). We chose to use existing visualizations in order to recreate conditions as close as possible to real-world scenarios. Specifically, we selected visualizations related to climate change due to prior work showing that such visualizations are likely to elicit high cognitive load in learners. First, climate visualizations can be hard to understand for non-experts [57], potentially eliciting a high level of intrinsic load. And, second, authors from the Intergovernmental Panel on Climate Change (IPCC) [33] often encounter challenges in overcoming visual complexity when designing their visualizations, which could lead to a high level of extraneous load. We chose to use *density maps* because we wanted to control the effect of participants' visualization literacy on extraneous cognitive load, and Lee *et al.* [48] found that similar representations were easy to decode for most people.

We selected two visualizations from Figure SPM.3 in the IPCC 6th Assessment Synthesis Report Summary for Policy Makers [19]. Figure SPM.3 (see the online version) presents three topics with multiple density maps each. We chose the first two topics because they had similar layouts, showed four maps each, and used the same colors, thus minimizing the possible effects of graphic design on extraneous load. Both images presented maps with projections across different scenarios of global warming for heat-humidity risks to human health (**Content A**) and species loss risk (**Content B**). In Content A, one of the maps depicted historical conditions, while the other three represented projections for different global warming ranges. Maps in content A only showed data for terrestrial parts of the world. In content B, all four maps showed projections at specific levels of global warming. Maps in content B showed data for both oceanic and terrestrial areas. To ensure good conceptual understandability for all participants regardless of their prior topic knowledge, we created short explanatory texts based on the IPCC Synthesis Report [19], IPCC's Working Group II full report, and their underlying studies. Each text first introduced the visualization's topic, explained the scientific mechanisms linking the data to climate risks, and briefly summarized projected trends over different global warming scenarios (we provide further details in App. A.1).

Presentation styles. we created two versions for each visualization: either *Compact* (style 1) or *Segmented* (style 2), as depicted in Fig. 1, and further detailed in Figs. 7 to 9 in App. A. The Compact style combined the data visual representation and text in a single image. The legend, title, and in-representation explanations mirrored those in the original IPCC report and we included our explanatory text below the image. In the Segmented style, we divided the same content into three images: (1) the visualization's title with the topic introduction part of our text, (2) the visualization's color legend along with the portion of our text explaining ecological or biological mechanisms at play, and (3) the visual representation, title, and legend, supplemented with the final part of our text describing trends across global warming scenarios. A slideshow container displayed these images, allowing survey respondents to control the pace of display by clicking on Previous / Next buttons (as shown in Fig. 10 in App. B), or using arrow keys on their keyboard. The content did not differ between styles, except for the legend, which appeared twice in the Segmented versions. All images had a fixed width of 800 pixels, ensuring identical sizes of texts and maps across all stimuli.

4.3 Measures

Cognitive load. We measured cognitive load through subjective mental effort ratings, as it has shown to correlate with other cognitive load measures such as electroencephalogram (EEG) [86], while being "unintrusive" [62] and easy to implement. Paas [60] first developed a 9-point Likert rating question that allowed learners to self-report mental effort invested in a learning task with responses ranging from *very, very low mental effort* to *very, very high mental effort.* Researchers have since adapted the wording in an effort to help learners disambiguate between different types of cognitive load in self-reports. Specifically, to assess **intrinsic load**, participants can rate how easy the learning content is from *very, very easy* to *very, very difficult* [4]. Similarly, to measure **extraneous load**, participants can report how easy it was to learn with the material from *very, very easy* to *very, very difficult* [22]. Finally, researchers suggested **germane load** could be measured by asking participants to rate how concentrated they were during learning from *very, very little* to *very, very much* [69, 86]. In this study, we adapted from these works (e. g., by replacing "learning content" [22] with "data content" for intrinsic load). To increase the reliability of collected data [25], we fully labeled all Likert answer options. In App. C.1, we provide the complete list of items and discuss their reliability.

Perceived readability. As we measured cognitive load from subjective ratings, we also collected subjective ratings of readability using PREVis [18], a validated questionnaire for evaluating perceived readability in data visualization. Respondents answered readability items on a 7-point bipolar agreement Likert scale, fully labeled. The instrument includes 11 items across four scales (see App. C.2 or the PREVis website for details):

◆ **UNDERSTAND**: how easily participants find they can *understand how to read* the visualization (three items, $\omega = 0.93$ for Style 1 and $\omega = 0.92$ for Style 2),

◆ **LAYOUT**: how clear they find the representation's *visual layout* (three items, $\omega = 0.79$ for Style 1 and $\omega = 0.92$ for Style 2),

◆ **DATAREAD**: how easily they feel they can *find and read data points* (three items, $\omega = 0.92$ for Style 1 and $\omega = 0.91$ for Style 2),

◆ **DATAFEAT**: how easily they feel they can *read data features* such as extremums, patterns, or trends (two items, $r = 0.70$ in Style 1, and $r = 0.53$ in Style 2).

We provide additional reliability analyses in App. C.2.

Participants answered each scale on a separate screen in random order, and items within each scale also appeared at random positions.

Style preference. At the end of the survey, participants indicated their style preference or their lack of preference for any style. They also provided a brief comment explaining their choice.

Data collection. We collected all measures electronically using

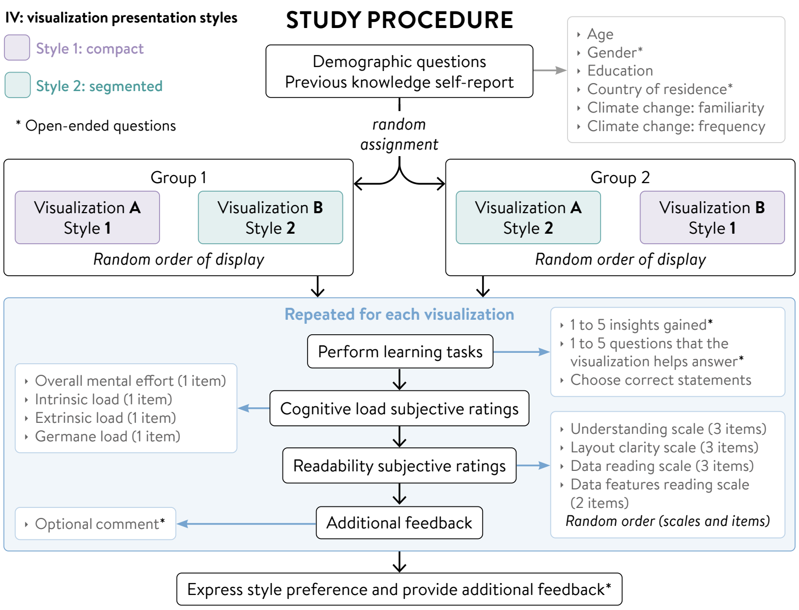


Figure 4: Overview of the study main part procedure.
Participants repeated the blue segment for each of the two visualizations they saw.

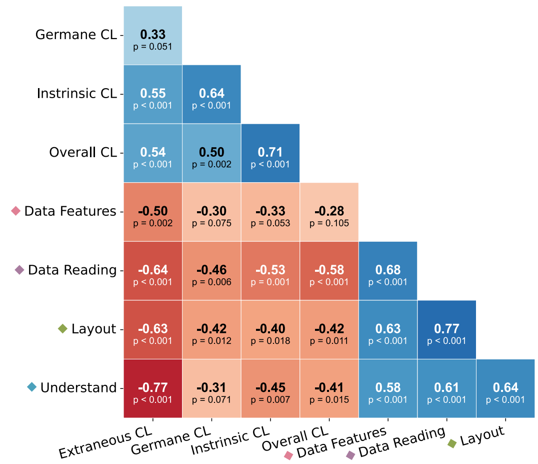


Figure 5: Repeated measures correlation matrix for subjective ratings of cognitive load (CL) and readability (PREVis ◆◆◆◆).

ou institutional LimeSurvey platform. We detail exclusion criteria in Appendix D.3 (e. g., mobile devices were not supported, and we included attention check questions).

4.4 Procedure

Each participant saw two visualizations: one was presented in three parts, according to the segmenting principle from CTML; and the other was presented as a single image. Both the content and order of appearance of these visualizations were randomized (see Figure 4). Due to participant dropout, however, the final number of complete submissions was uneven across conditions (see App. D.4). After clicking on the link in the invitation message, participants accessed the survey's welcome page, which stated again the requirements for entering the study. On the survey's next page, participants had to agree to an informed consent form, before they could proceed with the core part of the survey, which we describe in Fig. 4. Just before seeing the first visualization, the survey displayed an information page providing some context in the form of the first sentence from [19], the IPCC report from which we extracted the visualizations. We also included a brief message acknowledging that emotions arising in response to the climate crisis can be uncomfortable. We reminded participants that they could leave the survey anytime and provided an online list of resources to help cope with eco-anxiety. After participants had answered all questions about the first visualization, the survey recommended taking a short break before continuing with the final part of the study. On average, participants spent almost one hour on the survey ($M = 55$ min, $SD = 34$), removing one outlier participant whose recorded time was 545 min. Large discrepancies in recorded times for each question tend to indicate that some participants did not complete the survey in one go. A complete copy of the survey is available on `osf.io/cj3pe`

5 Results

5.1 Quantitative analyses

Table 1 describes the collected cognitive load and perceived readability ratings across conditions. 20 participants preferred the segmented style, six preferred the compact style, five had no preference, and three encountered a technical issue that caused missing data only for this very last question in the survey. We describe collected style preferences in App. D.5.

Readability and extraneous cognitive load correlation.
We created a repeated measures correlation matrix using the `rmcorr` package in `R` to examine the correlation between the perceived readability of visualizations and subjective extraneous cognitive load in the associated learning task. The resulting matrix in Figure 5 shows that readability ratings negatively correlated with extraneous load ratings (that is, a higher readability was associated with a lower extraneous cognitive load). To account for multiple comparisons between extraneous cognitive load ratings and PREVis ◆◆◆◆ scales ratings, we applied a Bonferroni correction, adjusting the significance threshold to .0125. The correlation was moderate for the ◆ DataFeat ratings ($r = -.50$, $p = .002$) and strong for the ◆ Understand ($r = -.77$, $p = 5.58 \times 10^{-8}$), ◆ Layout ($r = -.63$, $p = 4.49 \times 10^{-5}$), and ◆ DataRead ($r = -.64$, $p = 3.59 \times 10^{-5}$) ratings. These findings support our first hypothesis (H1): the more **readable** participants found a visualization, the less **extraneous** cognitive load they experienced (i. e., mental effort required to learn from the visualization). We provide details and describe other correlations as exploratory analyses in App. F.1.1.

Effect of segmentation on cognitive load. A paired-samples t-test showed that, although learners tended to experience a higher overall cognitive load in the compact condition (M = 5.59, SD = 1.5) compared to the segmented condition (M = 5.21, SD = 1.39), the difference did not reach the conventional significance threshold of .05 ($t(33) = 1.601$, $p = .060$). Since the results were marginally non-significant, and due to the small sample size for each randomized condition Group × Order, we did not conduct the other analyses planned in the pre-registration regarding this hypothesis. We did, however, conduct a series of exploratory paired-samples t-tests to assess the effect of presentation style on other cognitive load measures (see App. F.2.1). Fig. 6 shows the corresponding point estimates and confidence intervals. There was a significant difference for extraneous load between the compact and segmented styles, with a point estimate at −.88, but not for intrinsic and germane load.

As there were significant correlations between extraneous cognitive load and readability measurements, we also conducted a series of exploratory paired-samples t-tests to assess the effect of style on perceived readability (details in App. F.2.2). There were significant differences for the ◆ Understand, ◆ Layout, and ◆ DataRead scales, but not for the ◆ DataFeat scale (details in App. F.2.2).

Table 1: Collected ratings of cognitive load and readability for each Style × Content condition.

Random group	Group 1 ($N = 15$)				Group 2 ($N = 19$)			
Stimulus	Compact A		Segmented B		Segmented A		Compact B	
	mean	std	mean	std	mean	std	mean	std
Overall CL	5.80	1.37	5.27	1.28	5.16	1.50	5.42	1.61
Instrinsic CL	5.13	1.68	4.67	1.59	4.79	1.36	5.05	1.47
Extraneous CL	4.67	1.76	3.93	1.53	3.26	1.33	4.26	1.94
Germane CL	6.07	1.16	5.33	1.40	5.47	1.35	5.37	1.54
PREVis Understand	4.73	1.40	5.56	1.12	5.56	1.29	4.96	1.62
PREVis Layout	4.27	1.42	5.02	1.78	5.35	1.40	3.86	1.87
PREVis Data Features	4.90	1.71	5.43	1.49	5.11	1.52	4.74	1.65
PREVis Data Reading	4.38	1.51	4.96	1.27	5.39	1.34	4.72	1.72

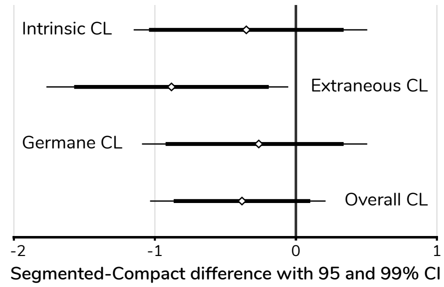


Figure 6: **Segmented**−**Compact** within-participants Cognitive Load (CL) differences.

5.2 Qualitative analysis of preference statements

At the end of the survey, participants expressed their preference in terms of presentation style and motivated their choice with a few sentences. We conducted a thematic analysis on the 29 comments we received. We provide the list of codes and themes in App. E and participants' coded answers in our `osf.io/pvrqh` repository.

This analysis revealed that many participants who preferred the segmented version felt that it guided them in progressively processing the information, and a few (P11, P13, P15) mentioned that the segmented style felt like a "story". In fact, P7 and P15 wrote about progressing from general context towards detailed information: "The segmented visualization felt like it told me the core takeways up front and then expanded into details" (P7). Such comments resonate with Shneiderman's ([76]) visual information-seeking mantra: "Overview first, zoom and filter, then details-on-demand". For participants who preferred the segmented version, the compact style felt "distracting" (P17, P23, P26, P28), "crowded" (P3, P6, P25, P28), or even "overwhelming" (P5, P17, P29), which could hinder their willingness to engage at all with the visualization: P5 mentioned they felt "discouraged by the look of the content". Some participants mentioned the compact presentation required more effort: "We have to find where to start and do this work ourself" (P15). The segmented version, in contrast, felt "clear" (P23, P26) or "appealing" (P9, P22) and required less effort as it was easier to "read" (P9, P18) and "understand" (P13, P14, P21). A portion of the participants, regardless of their preference, noted that having to go back and forth in the interactive version cost them time or was "annoying" (P4, P6, P12, P24, P29). Participants who either had no preference or preferred the compact version explained they found it better to see everything in one image (P1, P3, P24, P29). P24 and P30, who preferred the compact version, mentioned it made it easier to understand the content: "Having text and visualization in one place was easier for going back and forth to understand the topic better" (P24). It is worth noting that the segmented style was overall preferred and that the preferences were similarly distributed among self-reported levels of exposure to and familiarity with the topic (see Fig. 15).

6 Discussion

In our study, we observed patterns consistent with CLT-based predictions, with clear effects observed in extraneous load and perceived readability. These results, though exploratory, suggest that CLT and CTML can meaningfully inform how we conceptualize and assess cognitive processes in visualization. Our first hypothesis—that perceived readability ◆◆◆◆ would negatively correlate with subjective extraneous load ratings—was fully supported. Overall, this study's correlational findings add to existing empirical evidence supporting the relevance of CLT as a framework for evaluating data visualizations [3, 36, 50]. Our results specifically align with those of Huang *et al.* [36], who suggested linking visualization task performance to cognitive load. Further research should investigate the validity of other extraneous cognitive load measurements for evaluating readability—such as EEG measurements, secondary-task response time, or eye activity [26, 3, 63].

In addition, the ◆Understand readability scale showed a notably strong correlation with subjective ratings of extraneous cognitive load ($r = 0.77$), suggesting that participants' ability to comprehend *how to read* a data visualization greatly affected the amount of unnecessary cognitive load they experienced during learning tasks. This result supports the call from Stoiber *et al.* [77] to develop visualization *onboarding* practices, defined as "the process of supporting users on how to read, interpret, and extract information of visual representations of data." We further suggest that such onboarding procedures could benefit from integrating the CTML segmenting principle, which supports learners by presenting complex information in manageable chunks. Data visualization designers can implement segmentation through a progressive, scaffolded construction of data encodings—whether learners actively build the visualization themselves [38] or navigate through a pre-defined visual sequence, as in the present study. Some participants, however, found it cumbersome to navigate between segments in the visualization, suggesting that interactive features must be carefully designed to accommodate diverse preferences and avoid generating extraneous load. Non-interactive comics [87] also provide an relevant solution space.

As a second hypothesis (H2), we expected learners' overall mental effort to be lower in the segmented condition than in the compact condition. Although observed overall cognitive load differences approached conventional significance thresholds, we should not interpret them as evidence supporting H2. In their meta-analysis, Rey *et al.* [67] found that segmenting generally had a small effect on overall cognitive load, and no significant effect for learner-paced segmentation, as was our case. It is possible that the sample size in our current work was too small to detect a small effect, or that segmenting does not reduce overall cognitive load when learning from data visualizations. Another explanation relates to how germane cognitive load is thought to redistribute resources from extraneous activities to relevant information processing activities; in this view, segmentation would not affect the overall cognitive load, but would allow the learner to re-allocate resources from extraneous processing towards intrinsic load [82], thus improving the quality of learning. A study assessing learning gains—especially on topic understanding—through pre- and post-testing would help evaluate this hypothesis.

Beyond cognitive load redistribution, the expertise reversal effect in CTML [41, 42] may also account for the lack of significant effect of segmentation on overall cognitive load. This principle states that some design choices benefiting learners with low prior topic knowledge can have opposing effects on learners with high expertise—and vice-versa. Rey *et al.* [67] identified prior domain knowledge as a moderating factor in their meta-analysis, although they found effects contradictory to what the theory predicted. In a chapter on learner control, Scheiter [71] argues that only those with strong topic knowledge and clear guidance benefit from having control over their learning content and pacing. Novice learners, lacking the resources to make informed choices, may find this control to be a cognitive burden rather than a benefit. Unfortunately, our sample size was also too small to control for stimulus content and display order while examining potential interactions and effects on the overall cognitive load of style, self-reported topic familiarity, and topic exposure frequency. Besides, self-reports are not necessarily reliable indicators of objective prior knowledge. Further work could thus build on existing work [35, 68] to examine how the reversal expertise effect applies in data visualization, and investigate the mediating effect of objectively assessed domain knowledge on viewers' cognitive load.

To properly interpret these early findings, we must consider several limitations of the current study. First, the small sample size relative to the target population (general population, worldwide) limits the generalizability and precision of our findings. In addition, we lack reliable instruments for collecting subjective ratings of different types of cognitive load. As unreliable measuring instruments can significantly impede our ability to observe a phenomenon [27], a priority for future work is to develop and validate cognitive load assessment scales tailored to data visualization and other multimedia learning contexts, echoing recent calls to address the specificity of evaluation needs in visualization systems [85]. Finally, the study's online, unsupervised setting with unpaid participants also limited the reliability of cognitive load measurements and the usefulness of task time records (which we plot as additional information in App. G) as converging evidence of participants' cognitive engagement. Some participants may not have completed the survey in one go, which likely affected the results because subjective cognitive load ratings are highly sensitive to variations in mental workload [62]. Future work should aim to ensure more controlled experimental conditions to improve the accuracy of results. While our findings offer promising insights, additional research is required to address these limitations, and to identify possible mediating factors for data visualization-based knowledge acquisition, such as the viewer's prior knowledge or visualization literacy. As is often the case in both learning and data visualization contexts, no one-size-fits-all solution exists. Yet, our findings already indicate than CLT and CTML can serve as a *generative theoretical framework* [8] for evaluating and designing data visualizations.

7 Conclusion

In this work, we proposed an approach grounded in Cognitive Load Theory (CLT) and Cognitive Theory of Multimedia Learning (CTML) to design and evaluate data visualizations. We have only begun to explore the potential bridges between CLT and CTML, and visualization design and evaluation. In particular, many instructional design principles [54] remain untapped to shed light on, or to improve existing data visualization practices—for example: the animation principle for data-video design, the generative activity, mapping, and drawing principles for visualization educational activities, or the guided-discovery principle for system onboarding design. Visualization researchers might also consider how supporting metacognition and self-regulated learning strategies—like outlined in the metacognition in multimedia learning model and the cognitive load self-management principle—could assist even expert practitioners for complex visual analyses, or help them in learning how to use a new visual analytics system.

Visualization research is interdisciplinary by nature—just like cognitive science, and educational science. Visualization researchers thus have much to gain from borrowing conceptual frameworks from related fields such as cognitive psychology or educational science [74]. Initiatives like the EduVis workshop go a step further by building two-way bridges, notably through dual tracks—Ed4Vis and Vis4Ed—that connect both communities. We also see an opportunity for the data visualization community to contribute back to cognitive and learning sciences by providing a rich testing ground for theoretical development. In this sense, studies in visualization contexts could help us refine CTML and CLT: indeed, insights from more "applied" research domains, such as human-computer interaction, can raise new fundamental research questions for domains like cognitive or behavioral psychology. Visualization researchers, just like educational practitioners, can thus play an active role in advancing a *translational cognitive science* agenda [29], and help establish a dialogue between foundational research and applied practice.

Ultimately, by integrating knowledge from human-computer interaction research, educational science, and cognitive psychology, we can develop systems that better assist readers in comprehending information from data visualizations. This is particularly crucial when it comes to communicating about topics as significant as the risks associated with climate change.

Supplemental Materials

Figure Credits

Acknowledgments

The authors wish to thank all the participants in the study for their time and valuable feedback, as well as Macarena-Paz Celume and Nefeli Paparisteidi, who provided feedback on the work from a psychology and educational perspective.

References

[1] E. Adar and E. Lee. Communicative visualizations as a learning problem. *IEEE Trans Vis Comput Graph*, 27(2):946–956, 2021. doi: 10/ghgt48

[2] R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. In *Proc. InfoVis*, pp. 111–117. IEEE CS, Los Alamitos, 2005. doi: 10/bwrm27

[3] E. W. Anderson, K. C. Potter, L. E. Matzen, J. F. Shepherd, G. A. Preston, and C. T. Silva. A user study of visualization effectiveness using EEG and cognitive load. *Comput Graph Forum*, 30(3):791–800, 2011. doi: 10/cxcr9x

[4] P. Ayres. Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learn Instr*, 16(5):389–400, 2006. doi: 10/b5xhfj

[5] B. Bach, M. Keck, F. Rajabiyazdi, T. Losev, I. Meirelles, J. Dykes, R. S. Laramee, M. AlKadi, C. Stoiber, S. Huron, et al. Challenges and opportunities in data visualization education: A call to action. *IEEE Trans Vis Comput Graph*, 30(1):649–660, 2024. doi: 10/gtjwqj

[6] B. Bach, N. H. Riche, S. Carpendale, and H. Pfister. The emerging genre of data comics. *IEEE Comput Graph Appl*, 37(3):6–13, 2017. doi: 10/g9wq2s

[7] O. Barral, S. Lallé, A. Iranpour, and C. Conati. Effect of adaptive guidance and visualization literacy on gaze attentive behaviors and sequential patterns on magazine-style narrative visualizations. *ACM Trans Interactive Intell Syst*, 11(3–4), article no. 28, 46 pages, 2021. doi: 10/gtjwqm

[8] M. Beaudouin-Lafon, S. Bødker, and W. E. Mackay. Generative theories of interaction. *ACM Trans Comput-Hum Interact*, 28(6), article no. 45, 54 pages, 2021. doi: 10/pqrb

[9] R. Borgo, A. Abdul-Rahman, F. Mohamed, P. W. Grant, I. Reppa, L. Floridi, and M. Chen. An empirical study on using visual embellishments in visualization. *IEEE Trans Vis Comput Graph*, 18(12):2759–2768, 2012. doi: 10/f4fqd3

[10] M. A. Borkin, Z. Bylinskii, N. W. Kim, C. M. Bainbridge, C. S. Yeh, D. Borkin, H. Pfister, and A. Oliva. Beyond memorability: Visualization recognition and recall. *IEEE Trans Vis Comput Graph*, 22(1):519–528, 2016. doi: 10/ggf5r3

[11] M. A. Borkin, A. A. Vo, Z. Bylinskii, P. Isola, S. Sunkavalli, A. Oliva, and H. Pfister. What makes a visualization memorable? *IEEE Trans Vis Comput Graph*, 19(12):2306–2315, 2013. doi: 10/f5h3pd

[12] J. Boy, R. A. Rensink, E. Bertini, and J.-D. Fekete. A principled way of assessing visualization literacy. *IEEE Trans Vis Comput Graph*, 20(12):1963–1972, 2014. doi: 10/f6qjv6

[13] R. Brath. Metrics for effective information visualization. In *Proc. InfoVis*, pp. 108–111. IEEE CS, Los Alamitos, 1997. doi: 10/dm997r

[14] V. Braun and V. Clarke. Using thematic analysis in psychology. *Qual Res Psychol*, 3(2):77–101, 2006. doi: 10/fswdcx

[15] A. Burns, C. Lee, T. On, C. Xiong, E. Peck, and N. Mahyar. From invisible to visible: Impacts of metadata in communicative data visualization. *IEEE Trans Vis Comput Graph*, 30(7):3427–3443, 2024. doi: 10/gtx89f

[16] A. Burns, C. Xiong, S. Franconeri, A. Cairo, and N. Mahyar. How to evaluate data visualizations across different levels of understanding. In *Proc. BELIV*, pp. 19–28. IEEE CS, Los Alamitos, 2020. doi: 10/kz34

[17] A.-F. Cabouat, T. He, F. Cabric, T. Isenberg, and P. Isenberg. Position paper: A case to study the relationship between data visualization readability and visualization literacy. In *Proc. Visualization Literacy Workshop*, 2024. Online: `hal.science/hal-04523790`.

[18] A.-F. Cabouat, T. He, P. Isenberg, and T. Isenberg. PREVis: Perceived readability evaluation for visualizations. *IEEE Trans Vis Comput Graph*, 31(1):1083–1093, 2025. doi: 10/njnz

[19] K. Calvin, D. Dasgupta, G. Krinner, A. Mukherji, P. W. Thorne, C. Trisos, J. Romero, P. Aldunce, K. Barrett, G. Blanco, et al. IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Intergovernmental Panel on Climate Change (IPCC), Switzerland, 2023. doi: 10/k5jn

[20] L. Ciccione, M. Sablé-Meyer, E. Boissin, M. Josserand, C. Potier-Watkins, S. Caparos, and S. Dehaene. Trend judgment as a perceptual building block of graphicacy and mathematics, across age, education, and culture. *Sci Reports*, 13(1), article no. 10266, 11 pages, 2023. doi: 10/gtkckx

[21] L. Ciccione, M. Sablé-Meyer, and S. Dehaene. Analyzing the misperception of exponential growth in graphs. *Cognit*, 225, article no. 105112, 11 pages, 2022. doi: 10/g9r4ns

[22] G. Cierniak, P. Gerjets, and K. Scheiter. Expertise reversal in multimedia learning: Subjective load ratings and viewing behavior as cognitive process indicators. *Proc Annu Meet Cognit Sci Soc*, 31(31):1906–1911, 2009. Online: `escholarship.org/uc/item/99r5z5fb`.

[23] J. M. Clark and A. Paivio. Dual coding theory and education. *Educ Psychol Rev*, 3(3):149–210, 1991. doi: 10/c9gpvp

[24] F. R. Curcio. Comprehension of mathematical relationships expressed in graphs. *J Res Math Educ*, 18(5):382–393, 1987. doi: 10/fjh8jc

[25] A. DeCastellarnau. A classification of response scale characteristics that affect data quality: A literature review. *Qual Quant*, 52(4):1523–1559, 2018. doi: 10/gdqv89

[26] K. E. DeLeeuw and R. E. Mayer. A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *J Educ Psychol*, 100(1):223–234, 2008. doi: 10/cj4n4c

[27] R. F. DeVellis and C. T. Thorpe. *Scale Development: Theory and Applications*. SAGE Publications, 5th ed., 2021.

[28] M. Fansher, T. J. Adkins, P. Lalwani, A. Boduroglu, M. Carlson, M. Quirk, R. L. Lewis, P. Shah, H. Zhang, and J. Jonides. Icon arrays reduce concern over COVID-19 vaccine side effects: A randomized control study. *Cognit Res: Princ Implic*, 7(1), article no. 38, 9 pages, 2022. doi: 10/grgcmc

[29] B. Fisher, T. M. Green, and R. Arias-Hernández. Visual analytics as a translational cognitive science. *Top Cognit Sci*, 3(3):609–625, 2011. doi: 10/cc92x8

[30] A. R. Fox and J. D. Hollan. Visualization psychology: Foundations for an interdisciplinary research program. In D. Albers Szafir, R. Borgo, M. Chen, D. J. Edwards, B. Fisher, and L. Padilla, eds., *Visualization Psychology*, pp. 217–242. Springer, Cham, 2023. doi: 10/gtgz76

[31] M. Galesic and R. Garcia-Retamero. Graph literacy: A cross-cultural comparison. *Med Decis Making*, 31(3):444–457, 2011. doi: 10/fqxhjn

[32] M. Ghoniem, J.-D. Fekete, and P. Castagliola. On the readability of graphs using node-link and matrix-based representations: A controlled experiment and statistical analysis. *Inf Vis*, 4(2):114–135, 2005. doi: 10/bncqv6

[33] J. Harold, I. Lorenzoni, T. F. Shipley, and K. R. Coventry. Communication of IPCC visuals: IPCC authors' views and assessments of visual complexity. *Clim Change*, 158(2):255–270, 2020. doi: 10/gnffb6

[34] M. Hegarty. The cognitive science of visual-spatial displays: Implications for design. *Top Cognit Sci*, 3(3):446–474, 2011. doi: 10/c3274w

[35] B. D. Homer and J. L. Plass. Expertise reversal for iconic representations in science visualizations. *Instr Sci*, 38(3):259–276, 2010. doi: 10/dcfgcg

[36] W. Huang, P. Eades, and S.-H. Hong. Measuring effectiveness of graph visualizations: A cognitive load perspective. *Inf Vis*, 8(3):139–152, 2009. doi: 10/d3fs8b

[37] J. Hullman, E. Adar, and P. Shah. Benefitting InfoVis with visual difficulties. *IEEE Trans Vis Comput Graph*, 17(12):2213–2222, 2011. doi: 10/df5c8r

[38] S. Huron, S. Carpendale, A. Thudt, A. Tang, and M. Mauerer. Constructive visualization. In *Proc. DIS*, pp. 433–442. ACM, New York, 2014. doi: 10/gfspd2

[39] P. Isenberg. *Micro Visualizations*. HdR thesis, Université Paris-Saclay, France, 2021. Online: `hal.science/tel-03584024`.

[40] H. Jarodzka. Research methods in multimedia learning. In L. Fiorella and R. E. Mayer, eds., *The Cambridge Handbook of Multimedia Learning*, pp. 41–54. Cambridge UP, UK, 3rd ed., 2021. doi: 10/gtx89k

[41] S. Kalyuga. The expertise reversal principle in multimedia learning. In L. Fiorella and R. E. Mayer, eds., *The Cambridge Handbook of Multimedia Learning*, pp. 171–182. Cambridge UP, UK, 3rd ed., 2021. doi: 10/gtx893

[42] S. Kalyuga, P. Ayres, P. Chandler, and J. Sweller. The expertise reversal effect. *Educ Psychologist*, 38(1):23–31, 2003. doi: 10/d7rktj

[43] S. Kalyuga, P. Chandler, and J. Sweller. Incorporating learner experience into the design of multimedia instruction. *J Educ Psychol*, 92(1):126–136, 2000. doi: 10/d8s9q3

[44] J. Krosnick. Questionnaire design. In *The Palgrave Handbook of Survey Research*, pp. 439–455. Palgrave Macmillan, Cham, 2017. doi: 10/gqcw3j

[45] B. C. Kwon and B. Lee. A comparative evaluation on online learning approaches using parallel coordinate visualization. In *Proc. CHI*, pp. 993–997. ACM, New York, 2016. doi: 10/ghppzx

[46] J. H. Larkin and H. A. Simon. Why a diagram is (sometimes) worth ten thousand words. *Cognit Sci*, 11(1):65–100, 1987. doi: 10/d9pkk4

[47] S. Lee, S.-H. Kim, Y.-H. Hung, H. Lam, Y.-A. Kang, and J. S. Yi. How do people make sense of unfamiliar visualizations?: A grounded model of novice's information visualization sensemaking. *IEEE Trans Vis Comput Graph*, 22(1):499–508, 2016. doi: 10/gfw4vs

[48] S. Lee, S.-H. Kim, and B. C. Kwon. VLAT: Development of a visualization literacy assessment test. *IEEE Trans Vis Comput Graph*, 23(1):551–560, 2017. doi: 10/f92d38

[49] E. Lee-Robbins, S. He, and E. Adar. Learning objectives, insights, and assessments: How specification formats impact design. *IEEE Trans Vis Comput Graph*, 28(1):676–685, 2022. doi: 10/gtx89j

[50] P. D. Mautone and R. E. Mayer. Cognitive aids for guiding graph comprehension. *J Educ Psychol*, 99(3):640–652, 2007. doi: 10/dc3dcr

[51] R. E. Mayer, ed. *Multimedia Learning*. Cambridge UP, UK, 2001. doi: 10/gtx89q

[52] R. E. Mayer. Multimedia principle. In *Multimedia Learning*, pp. 63–80. Cambridge UP, UK, 2001. doi: 10/gtx89r

[53] R. E. Mayer. Cognitive theory of multimedia learning. In L. Fiorella and R. E. Mayer, eds., *The Cambridge Handbook of Multimedia Learning*, pp. 57–72. Cambridge UP, UK, 3rd ed., 2021. doi: 10/gtx89t

[54] R. E. Mayer and L. Fiorella, eds. *The Cambridge Handbook of Multimedia Learning*. Cambridge UP, UK, 3rd ed., 2021. doi: 10/gtx89s

[55] R. E. Mayer and L. Fiorella. Principles for managing essential processing in multimedia learning: Segmenting, pre-training, and modality principles. In R. E. Mayer and L. Fiorella, eds., *The Cambridge Hand-*

book of Multimedia Learning, pp. 243–260. Cambridge UP, UK, 3rd ed., 2021. doi: 10/gtx89z
[56] M. T. McCrudden and P. N. Van Meter. Multimedia learning with visual displays. In L. Fiorella and R. E. Mayer, eds., *The Cambridge Handbook of Multimedia Learning*, pp. 510–520. Cambridge UP, UK, 3rd ed., 2021. doi: 10/gtx89w
[57] R. McMahon, M. Stauffacher, and R. Knutti. The unseen uncertainties in climate change: Reviewing comprehension of an IPCC scenario graph. *Clim Change*, 133(2):141–154, 2015. doi: 10/f7v2bq
[58] M. E. Milesi, P. Mejia-Domenzain, L. Brandl, V. Echeverria, Y. Jin, D. Gasevic, Y.-S. Tsai, T. Käser, and R. Martinez-Maldonado. "piecing data connections together like a puzzle": Effects of increasing task complexity on the effectiveness of data storytelling enhanced visualisations. In *Proc. CHI*, article no. 618, 18 pages. ACM, New York, 2025.
[59] T. Munzner. *Visualization Analysis and Design*. CRC Press, Boca Raton, 2014. doi: 10/gd3xgq
[60] F. Paas. Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *J Educ Psychol*, 84(4):429–434, 1992. doi: 10/c74hkc
[61] F. Paas and J. Sweller. Implications of cognitive load theory for multimedia learning. In L. Fiorella and R. E. Mayer, eds., *The Cambridge Handbook of Multimedia Learning*, pp. 73–81. Cambridge UP, UK, 3rd ed., 2021. doi: 10/gtx89n
[62] F. Paas, J. E. Tuovinen, H. Tabbers, and P. W. M. Van Gerven. Cognitive load measurement as a means to advance cognitive load theory. *Educ Psychologist*, 38(1):63–71, 2003. doi: 10/bznn3t
[63] L. M. Padilla, S. C. Castro, P. S. Quinan, I. T. Ruginski, and S. H. Creem-Regehr. Toward objective evaluation of working memory in visualizations: A case study using pupillometry and a dual-task paradigm. *IEEE Trans Vis Comput Graph*, 26(1):332–342, 2020. doi: 10/ghs95w
[64] A. Paivio. Mental imagery in associative learning and memory. *Psychol Rev*, 76(3):241–263, 1969. doi: 10/d8xprn
[65] G. J. Quadri and P. Rosen. A survey of perception-based visualization studies by task. *IEEE Trans Vis Comput Graph*, 28(12):5026–5048, 2022. doi: 10/gr6323
[66] G. J. Quadri, A. Z. Wang, Z. Wang, J. Adorno, P. Rosen, and D. A. Szafir. Do You See What I See? A Qualitative Study Eliciting High-Level Visualization Comprehension. In *Proc. CHI*, article no. 204, 26 pages. ACM, New York, 2024. doi: 10/gtx892
[67] G. D. Rey, M. Beege, S. Nebel, M. Wirzberger, T. H. Schmitt, and S. Schneider. A meta-analysis of the segmenting effect. *Educ Psychol Rev*, 31(2):389–419, 2019. doi: 10/gtx89x
[68] J. Richter, A. Wehrle, and K. Scheiter. How the poor get richer: Signaling guides attention and fosters learning from text-graph combinations for students with low, but not high prior knowledge. *Appl Cognit Psychol*, 35(3):632–645, 2021. doi: 10/gndvdd
[69] G. Salomon. Television is "easy" and print is "tough": The differential investment of mental effort in learning as a function of perceptions and attributions. *J Educ Psychol*, 76(4):647–658, 1984. doi: 10/bq9z6d
[70] P. Sánchez-Holgado, C. Arcila-Calderón, and M. Frías-Vázquez. The effect of interest and attitude on public comprehension of news with data visualization. *Front Commun*, 8(9), article no. 1064184, 9 pages, 2023. doi: 10/gtx89g
[71] K. Scheiter. The learner control principle in multimedia learning. In L. Fiorella and R. E. Mayer, eds., *The Cambridge Handbook of Multimedia Learning*, pp. 418–429. Cambridge UP, Cambridge, 3rd ed., 2021. doi: 10/g9r4nq
[72] A. Schlieder, J. Rummel, P. Albers, and F. Sadlo. Sequential visual cues from gaze patterns: Reasoning assistance for bar charts. In *Proc. CHI*, article no. 621, 17 pages. ACM, New York, 2025. doi: 10/g9pv62
[73] W. Schnotz and T. Rasch. Enabling, facilitating, and inhibiting effects of animations in multimedia learning: Why reduction of cognitive load can have negative results on learning. *Educ Technol Res Dev*, 53(3):47–58, 2005. doi: 10/fswbxk
[74] K. Schönborn and L. Besançon. What can educational science offer visualization? a reflective essay. In *Proc. EduVIS*, pp. 30–37. doi: 10/g9p2r8
[75] P. Shah and J. Hoeffner. Review of graph comprehension research: Implications for instruction. *Educ Psychol Rev*, 14(1):47–69, 2002. doi: 10/cq7tq5
[76] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proc. VL*, pp. 336–343. IEEE CS, Los Alamitos, 1996. doi: 10/fwdq26
[77] C. Stoiber, F. Grassinger, M. Pohl, H. Stitz, M. Streit, and W. Aigner. Visualization onboarding: Learning how to read and use visualizations. OSF preprints, 2019. doi: 10/gh38zd
[78] C. Stoiber, M. Wagner, F. Grassinger, M. Pohl, H. Stitz, M. Streit, B. Potzmann, and W. Aigner. Visualization onboarding grounded in educational theories. In D. Albers Szafir, R. Borgo, M. Chen, D. J. Edwards, B. Fisher, and L. Padilla, eds., *Visualization Psychology*, pp. 139–164. Springer, Cham, 2023. doi: 10/gtjwqn
[79] C. Stoiber, C. Walchshofer, M. Pohl, B. Potzmann, F. Grassinger, H. Stitz, M. Streit, and W. Aigner. Comparative evaluations of visualization onboarding methods. *Vis Inf*, 6(4):34–50, 2022. doi: 10/g9r4nr
[80] J. Sweller. Cognitive load during problem solving: Effects on learning. *Cognit Sci*, 12(2):257–285, 1988. doi: 10/cmgsv8
[81] J. Sweller. Cognitive load theory and educational technology. *Educ Technol Res Dev*, 68(1):1–16, 2020. doi: 10/ggjgpr
[82] J. Sweller, J. J. G. van Merriënboer, and F. Paas. Cognitive architecture and instructional design: 20 years later. *Educ Psychol Rev*, 31(2):261–292, 2019. doi: 10/gftwkk
[83] A. Thudt, J. Walny, C. Perin, F. Rajabiyazdi, L. MacDonald, R. Vardeleon, S. Greenberg, and S. Carpendale. Assessing the readability of stacked graphs. In *Proc. GI*, pp. 167–174. Canadian Human-Computer Communications Society, Mississauga, 2016. doi: 10/gtgz8d
[84] T. van Gog. The Signaling (or Cueing) Principle in Multimedia Learning. In R. E. Mayer, ed., *The Cambridge Handbook of Multimedia Learning*, pp. 263–278. Cambridge UP, UK, 2nd ed., 2014. doi: 10/g9pv6z
[85] E. Z. Victorelli, A.-F. Cabouat, E. Santos, F. Cabric, and P. Isenberg. We should change how we measure user experience in visual analytics systems. In *Proc. EuroVA*. Eurographics Assoc, Goslar, 2025. doi: 10/g9wq2t
[86] J. Wang, P. Antonenko, A. Keil, and K. Dawson. Converging subjective and psychophysiological measures of cognitive load to study the effects of instructor-present video. *Mind, Brain, Educ*, 14(3):279–291, 2020. doi: 10/ggwqgp
[87] Z. Wang, L. Sundin, D. Murray-Rust, and B. Bach. Cheat sheets for data visualization techniques. In *Proc. CHI*, article no. 144, 13 pages. ACM, New York, 2020. doi: 10/gnrq7t
[88] Y. L. Wong, K. Madhavan, and N. Elmqvist. Towards characterizing domain experts as a user group. In *Proc. BELIV*, pp. 1–10. IEEE CS, Los Alamitos, 2018. doi: 10/gtx89h
[89] V. Yoghourdjian, Y. Yang, T. Dwyer, L. Lawrence, M. Wybrow, and K. Marriott. Scalability of network visualisation from a cognitive load perspective. *IEEE Trans Vis Comput Graph*, 27(2):1677–1687, 2021. doi: 10/ghs96j
[90] C. Ziemkiewicz, A. Ottley, R. J. Crouser, K. Chauncey, S. L. Su, and R. Chang. Understanding visualization by understanding individual users. *IEEE Comput Graph Appl*, 32(6):88–94, 2012. doi: 10/gg9qjw

Summary of Appendices

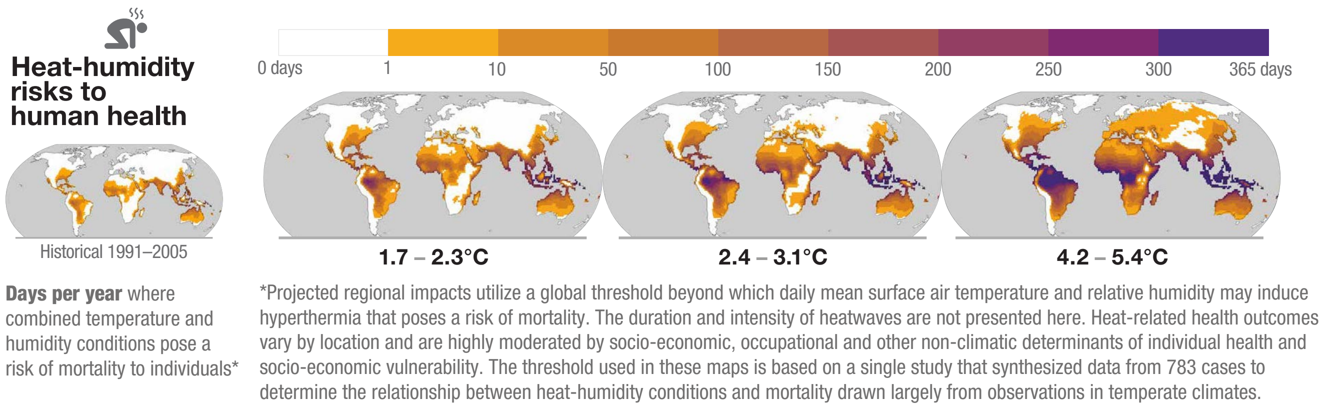
A Stimuli design: figures and accompanying text

As described in Sec. 4.2, we based our stimuli on two visualizations from Figure SPM.3 in the IPCC 6th Assessment Synthesis Report Summary for Policy Makers [19]. We found it would be hard for participants to fully understand the content once taken out of the context of the Summary Report. Additionally, the provided captions in the Summary Report were quite technical (see below in App. A.1.1 and App. A.1.2). We thus replaced the original captions with an explanatory text based on the full report from IPCC Working Group II, and the relevant underlying studies, as detailed below.

A.1 Compact Style (Style 1)

Figs. 7 and 8 below present the study's stimuli contents in the Compact style (Style 1). Upper parts of the figures correspond to the original representations from parts *a)* and *b)* of Figure SPM3 in [19]. The text block below corresponds to the explanatory text we created, which in the Compact style resembled a caption below the visualization.

A.1.1 Content A: Heat-humidity risk to human health



The visualization above presents climate change data related to the risk of heat-humidity conditions that exceed human thermoregulatory capacity. Heat-humidity risk is not only related to temperature: the boundary at which the environment's temperature becomes deadly decreases with increasing relative humidity (Mora et al., 2017). If the external temperature is too hot, the body must sweat (i.e., perspiration) to regulate its internal temperature. However, if the air is too humid, the sweat cannot evaporate, and the body's temperature cannot decrease. The more days of exposure to such conditions, the more likely humans are to die from hyperthermia or to suffer heat-related metabolic stress, leading to increased morbidity. Different climate change scenarios relate to varying levels of risk in the future; the visualization above shows projections for the year 2100 according to 3 different global warming scenarios. Without rapid, deep, and sustained mitigation and accelerated adaptation actions, losses and damages will continue to increase, and they will disproportionately affect the most vulnerable populations.

Figure 7: Content A presented in Style 1 (compact).

The upper part is identical to the original visualization SPM3 part b). The original associated caption in the Summary Report reads as follows:

> "Risks to human health as indicated by the days per year of population exposure to hyperthermic conditions that pose a risk of mortality from surface air temperature and humidity conditions for historical period (1991–2005) and at GWLs of 1.7°C–2.3°C (mean = 1.9°C; 13 climate models), 2.4°C–3.1°C (2.7°C; 16 climate models) and 4.2°C–5.4°C (4.7°C; 15 climate models). Interquartile ranges of GWLs by 2081–2100 under RCP2.6, RCP4.5 and RCP8.5. The presented index is consistent with common features found in many indices included within WGI and WGII assessments. "

We replaced these technical specifications with a text composed of three parts:

1. **An explanation of the physical and physiological mechanisms at play for this risk:**

   > "The visualization above presents climate change data related to the risk of heat-humidity conditions that exceed human thermoregulatory capacity. Heat-humidity risk is not only related to temperature: the boundary at which the environment's temperature becomes deadly decreases with increasing relative humidity (Mora et al., 2017). If the external temperature is too hot, the body must sweat (i.e., perspiration) to regulate its internal temperature. However, if the air is too humid, the sweat cannot evaporate, and the body's temperature cannot decrease."

2. **An explanation of the key indicator chosen in the visualization**, in this case, the number of days of exposure per year, which was encoded as a binned, linear color scale:

   > "The more days of exposure to such conditions, the more likely humans are to die from hyperthermia or to suffer heat-related metabolic stress, leading to increased morbidity."

3. **A overview-level takeaway on the topic:**

   > "Different climate change scenarios relate to varying levels of risk in the future; the visualization above shows projections for the year 2100 according to 3 different global warming scenarios. Without rapid, deep, and sustained mitigation and accelerated adaptation actions, losses and damages will continue to increase, and they will disproportionately affect the most vulnerable populations."

To develop this explanatory text, we consulted relevant sections of the IPCC Working Group II report: "Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change." doi:10.1017/9781009325844. Much of the information on heat-humidity presented in the full report—and underlying the Summary Report visualization—derived from a single study: C. Mora *et al.*, "Global risk of deadly heat," Nature Clim Change, vol. 7, no. 7, pp. 501–506, 2017, doi: 10.1038/nclimate3322. We therefore also reviewed this study to inform our written content and ensure scientific accuracy.

A.1.2 Content B: Risk of species losses

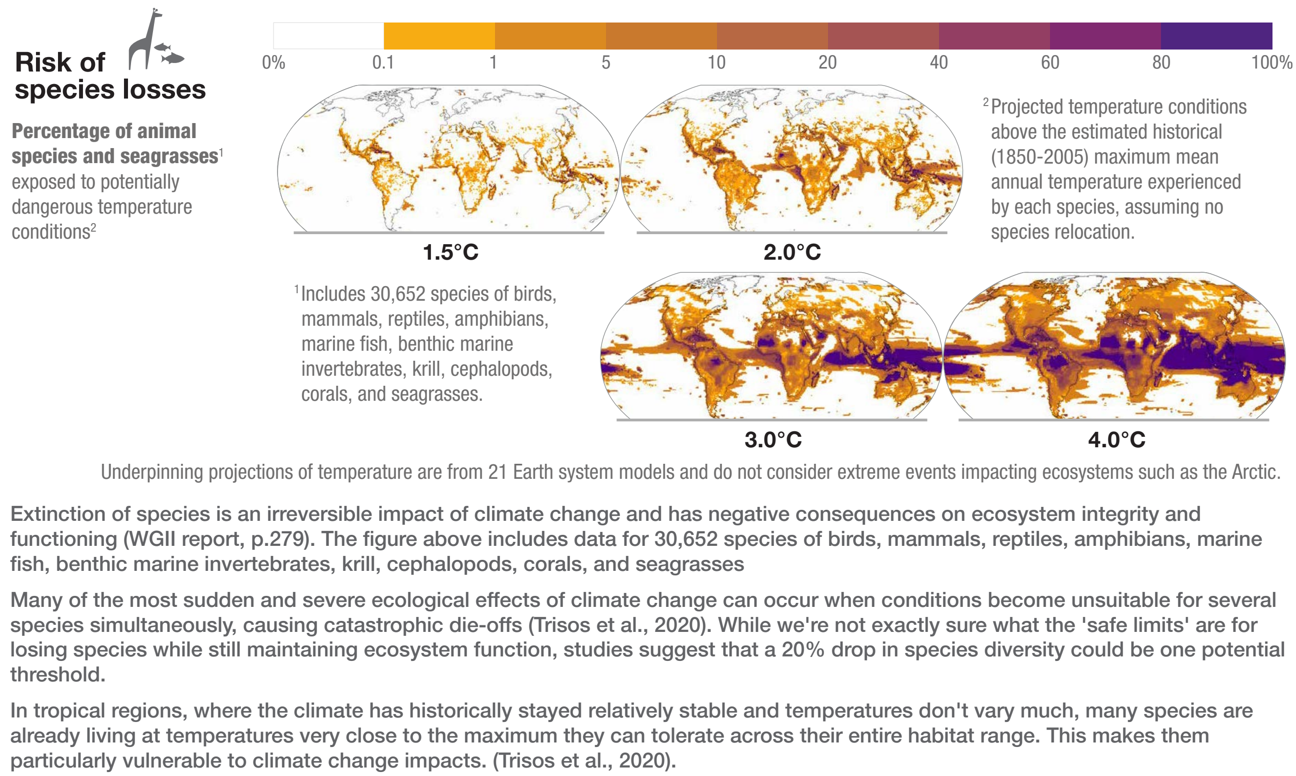


Figure 8: Content A presented in Style 1 (compact).

The upper part is identical to the original visualization SPM3 part a). The original associated caption in the Summary Report read as follows:

> "Risks of species losses as indicated by the percentage of assessed species exposed to potentially dangerous temperature conditions, as defined by conditions beyond the estimated historical (1850–2005) maximum mean annual temperature experienced by each species, at GWLs of 1.5°C, 2°C, 3°C and 4°C. Underpinning projections of temperature are from 21 Earth system models and do not consider extreme events impacting ecosystems such as the Arctic."

We replaced these technical specifications with a text composed of three parts:

1. **An explanation of the ecological mechanisms at play for this risk:**

   > "Extinction of species is an irreversible impact of climate change and has negative consequences on ecosystem integrity and functioning (WGII report, p.279). The figure above includes data for 30,652 species of birds, mammals, reptiles, amphibians, marine fish, benthic marine invertebrates, krill, cephalopods, corals, and seagrasses."

2. **An explanation of the key indicator chosen in the visualization**, in this case, the percentage of species losses, on a binned, non-linear scale centered around 20%:

   > "Many of the most sudden and severe ecological effects of climate change can occur when conditions become unsuitable for several species simultaneously, causing catastrophic die-offs (Trisos et al., 2020). While we're not exactly sure what the 'safe limits' are for losing species while still maintaining ecosystem function, studies suggest that a 20% drop in species diversity could be one potential threshold."

3. **A overview-level takeaway on the topic:**

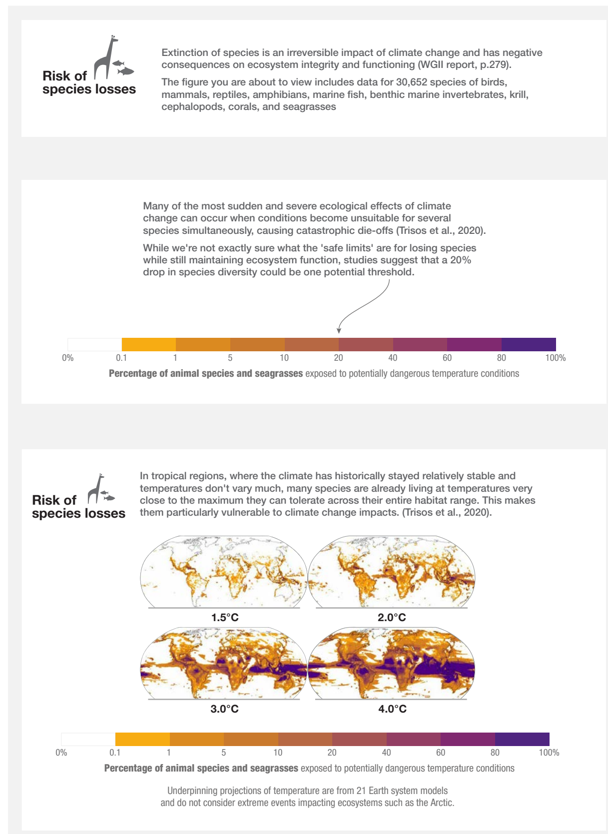
   > "In tropical regions, where the climate has historically stayed relatively stable and temperatures don't vary much, many species are already living at temperatures very close to the maximum they can tolerate across their entire habitat range. This makes them particularly vulnerable to climate change impacts. (Trisos et al., 2020)."

To develop this explanatory text, we consulted relevant sections of the IPCC Working Group II report: "Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change." doi:10.1017/9781009325844. Much of the information on the risk of species losses presented in the full report—and underlying the Summary Report visualization—derived from a single study: C. H. Trisos, C. Merow, and A. L. Pigot, "The projected timing of abrupt ecological disruption from climate change," Nature, vol. 580, no. 7804, pp. 496–501, 2020, doi: 10.1038/s41586-020-2189-9. We therefore also reviewed this study to inform our written content and ensure scientific accuracy.

A.2 Segmented Style (Style 2)

The content of stimuli in the Segmented style remained the same as in the Compact style (Style 1, Figs. 7 and 8) but was segmented as described in the Method subsection 4.2. The main difference was that the legend was first introduced in the second segment along with part 2 of the explanatory text we created (see details in App. A.1). The legend was thus presented twice. The interface for navigating between segments is presented in App. B Fig. 10.

Content A presented in Style 2.



Content B presented in Style 2.

Figure 9: Segmented versions of the visualizations (Style 2). Source: Figure SPM3 in [19].

B Survey screenshot (segmented visualization condition)

Figure 10: Example screenshot for the presentation of visualization B2, i. e., content B (risk of species loss) in style 2 (Segmented). The image in the figure shows the second segment out of three. The screenshot also shows the survey's accompanying text above the visualization and up/down buttons available to navigate between visualization segments in the Segmented condition. In this condition, participants needed to see each segment at least once before being allowed to move on to the next page in the survey.

C MEASURING INSTRUMENTS

C.1 Cognitive load

Cognitive load dimension	Item question	9-point Likert labeling
Overall load	How much mental effort did you invest to complete the task?	From *very, very low mental effort* to *very, very high mental effort*
Instrinsic load	How simple was the data content?	From *very, very simple* to *very, very complex*
Extraneous load	How easy was it to use the data visualization for completing the task?	From *very, very simple* to *very, very complex*
Germane load	How focused were you during task completion?	From *very, very little* to *very, very much*

Table 2: Questions and Likert extremum labels used to collect subjective cognitive load ratings. All points were numbered and labeled, as shown in Figure 11 above.

Since there are no standardized questionnaires to measure the different types of cognitive load components from Cognitive Load Theory , we based our items for assessing subjective cognitive load on previous works—as we describe in Sec. 4.3. Tab. 2 above shows the final items (question and answer option labels) we implemented in our survey. We adapted from the original items in the literature as follows:

- **Overall cognitive load:** the original question [60] asked learners how much mental effort they invested "in the problem". Since our learning task did not involve solving a problem, we adapted the question with the words "to complete the task".
- **Intrinsic cognitive load:** similarly, the original question was created in a problem-solving context [4] and asked learners to rate "how easy or difficult they found each calculation" on a 7-point scale ranging from 1 (extremely easy) to 7 (extremely difficult). In subsequent research [22], researchers used a different question: "How difficult was the learning content for you?". Participants responded using a six-point Likert-type scale ranging from "not at all" (1 point) to "extremely" (6 points).
  We adapted our question to help learners focus on the data underlying the visualization and asked "How simple was the data content?" To increase the reliability of the ratings across all four subjective cognitive load ratings, we expanded the answer options into a 9-point Likert scale ranging from 1 (very, very simple) to 9 (very, very complex).
- **Extraneous cognitive load:** The original question used to evaluate subjective extraneous load read asked "How difficult was it for you to learn with the material?" on a six-point Likert-type scale ranging from "not at all" (1 point) to "extremely" (6 points) [22]. Because we did not frame our questions as an educational intervention, we reworded as follow: "How easy was it to use the data visualization for completing the task?" and aligned the answer options to a 9-point Likert scale.
- **Germane cognitive load:** Similarly to extraneous load, the original germane cognitive load question asked participants to rate on a 4-point Likert scale how much they concentrated during "reading" or "watching", depending on the experimental condition [69]. In subsequent work [22], researchers replaced the term with "learning" and used a six-point Likert-type scale ranging from "not at all" (1 point) to "extremely" (6 points). We asked our participants how focused they were "during task completion" and expanded the answer options to match the other questions' 9-point Likert scale.

As a final adjustment, to increase the reliability of collected data [25], we fully labeled all Likert answer options, as described in Fig. 11.

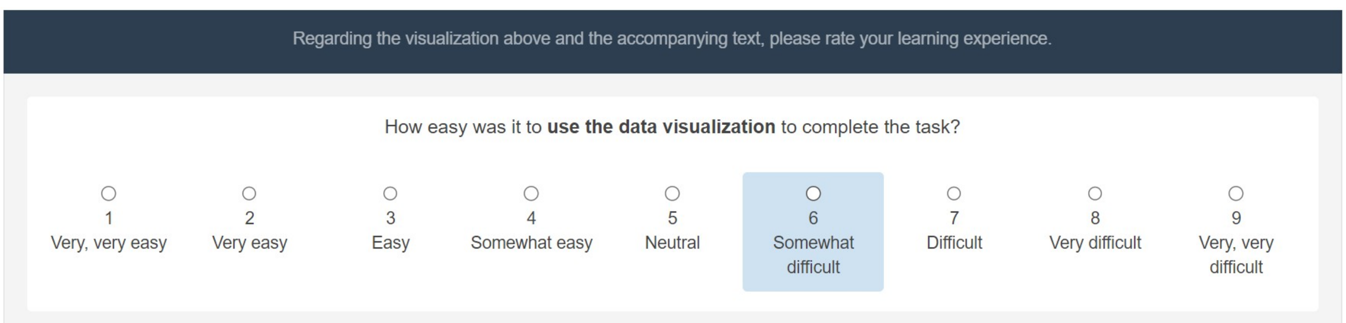


Figure 11: Screenshot of the "Extraneous load" rating item and answer options.

Because all four questions targeted related aspects of mental effort (overall, intrinsic, extraneous, and germane cognitive load) but weren't standardized items, we examined internal consistency as a reliability check. In the Compact (Style 1) condition, the group of questions demonstrated good internal consistency ($\alpha = 0.81$ and $\omega = 0.86$). In the Segmented (Style 2) condition, internal consistency was somewhat lower but still acceptable ($\alpha = 0.7$ and $\omega = 0.78$).

To explore the lower internal consistency observed in the Segmented condition (Style 2), we inspected inter-item correlations for each style (see Fig. 12). In Style 1, cognitive load questions showed moderate to strong positive correlations, indicating that participants responded similarly across questions targeted at different types of mental effort. In contrast, correlations were notably weaker in Style 2, suggesting that participants more clearly distinguished between aspects of cognitive load—possibly reflecting changes in how cognitive resources were allocated under the two visualization styles.

In particular, Extrinsic load showed such weaker correlations with Overall load in the Segmented condition ($r = .17$) than in the Compact condition ($r = .53$), suggesting that participants' perceptions of the mental effort required to process the visualization were less associated with their overall sense of mental effort. This may indicate that the segmented format reduced the perceived impact of representational processing demands on overall task difficulty. Similarly, Extrinsic load showed much weaker correlations with Intrinsic load in the Segmented condition ($r = .28$) compared to the Compact condition ($r = .68$), suggesting that the Segmented visualization may have helped participants to isolate essential content from design-induced difficulties, resulting in a clearer separation between different types of cognitive load.

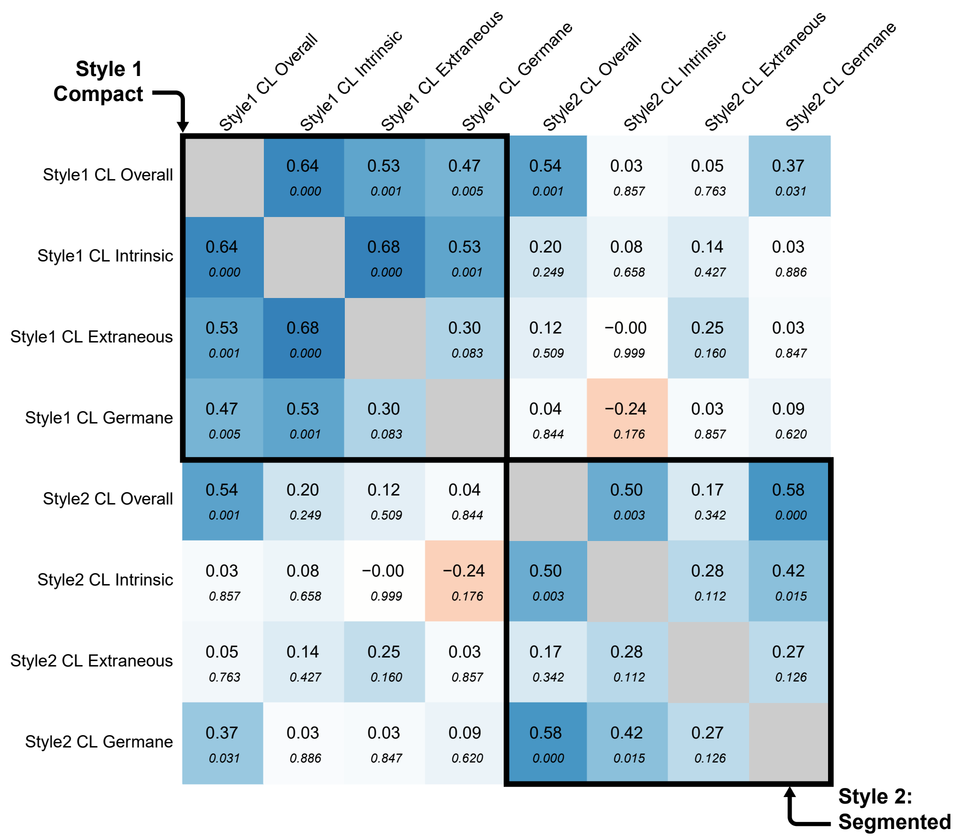


Figure 12: Inter-items correlations for cognitive load subjective ratings for each stimulus style (Compact and Segmented).

C.2 Perceived readability (PREVis)

Table 3: PREVis rating scales, items and answer options.

Scale	Item code	Answer options
◆ UNDERSTAND	It is obvious for me how to read this visualization	*1 - Strongly disagree*
	I can easily understand how the data is represented in this visualization	*2 - Disagree*
	I can easily understand this visualization	*3 - Slightly disagree*
◆ LAYOUT	I don't find this visualization messy	*4 - Neutral*
	I don't find this visualization crowded	*5 - Slightly agree*
	I don't find parts of the visualization distracting	*6 - Agree*
◆ DATAFEAT	I find data features (for example, a minimum, or an outlier, or a trend) visible in this visualization	*7 - Strongly agree*
	I can clearly see data features (for example, a minimum, or an outlier, or a trend) in this visualization	
◆ DATAREAD	I can easily retrieve information from this visualization	
	I can easily identify relevant information in this visualization	
	I can easily find specific elements in this visualization	

Table 4: Internal consistency for PREVis scales in our survey ($N = 34$). For scales with 3 items, we calculate Cronbach's α and McDonald's ω. The ◆ **DataFeat** scale has only two items: in this case, Cronbach's α is mathematically equivalent to Pearson's r and we cannot calculate MacDonald's ω.

Subscale	Stimulus style	α	ω	Pearson's r
◆ **Understand**	Style 1	0.92	0.93	–
	Style 2	0.91	0.92	–
◆ **Layout**	Style 1	0.79	0.79	–
	Style 2	0.91	0.92	–
◆ **DataFeat**	Style 1	–	–	0.70
	Style 2	–	–	0.53
◆ **DataRead**	Style 1	0.91	0.92	–
	Style 2	0.87	0.91	–

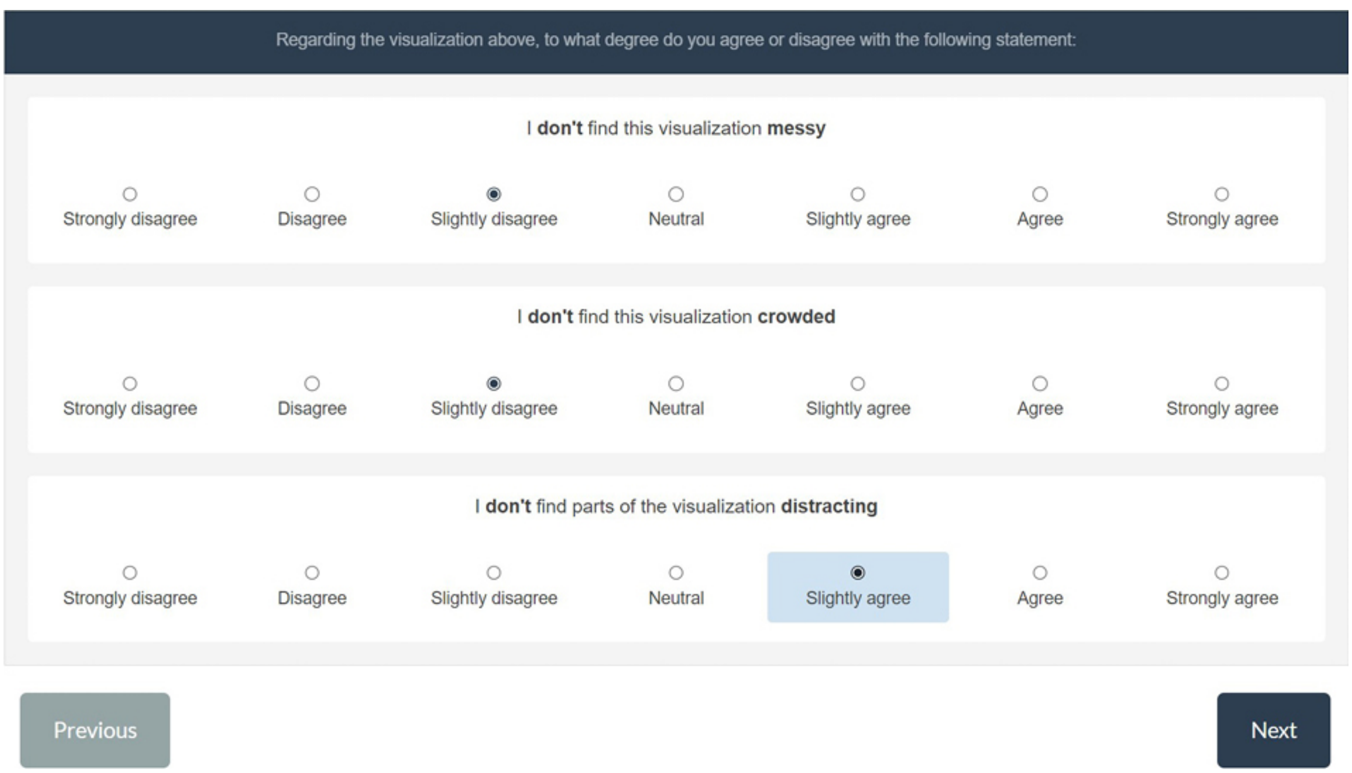


Figure 13: Example screenshot of the "Layout" PREVis scale.

D Description of collected data

D.1 Demographics table

Table 5: Demographics of participants (categories without participants are omitted).

		N
Age	18-24	3
	25-34	22
	35-44	4
	45-54	1
	55-64	2
	65+	2
Gender	Man	18
	Non-binary	1
	Woman	15
Education	Bachelor	9
	Master	21
	PhD	3
	Secondary	1
Country	Australia	1
	Austria	2
	France	16
	Germany	4
	Kenya	3
	Netherlands	1
	Pakistan	2
	Saudi Arabia	3
	USA	2

D.2 Previous exposure to climate change

Table 6: Count of participants' answers for the climate change familiarity self-report: "*How familiar are you with the topic of climate change science?*".

Likert value	Description	*N*
1	I don't know what climate change is.	0
2	—	3
3	—	11
4	I actively follow the discussion in the media on climate change.	14
5	—	3
6	—	3
7	I am a climate scientist myself.	0

Table 7: Count of participants' answers for the climate change information frequency self-report: "*How often do you interact with documents or media related to climate change science?*"

Likert value	Description	*N*
1	I have never read or watched any content related to climate change in my life.	0
2	I rarely (a few times a year) read or watch content related to climate change in my life.	6
3	I sometimes (1-3 times a month) read or watch content related to climate change, such as in news articles.	16
4	I frequently (once a week or more) read or watch content related to climate change, such as in news articles.	8
5	I sometimes (a few times a year) create content related to climate change, either for work or as a hobby.	3
6	I frequently (several times per month) create content related to climate change, either for work or as a hobby.	0
7	Studying climate change and/or creating related content represents the core of my daily activities as a professional.	1

D.3 Data exclusion

D.3.1 Incomplete answers

Out of 475 people who accessed the online survey, 425 of them didn't proceed past the first page. A reason for this high dropout rate could be that, even though we specified in the invitation that participants needed to access the survey from a desktop computer, many people may have clicked on the link without reading this information. In order to highlight this requirement, we had included the image from Figure 14 on the first page. As a result, many individuals may have dropped off from the first page because they tried to open the survey on their smartphones, only to then realize we required a desktop setting.

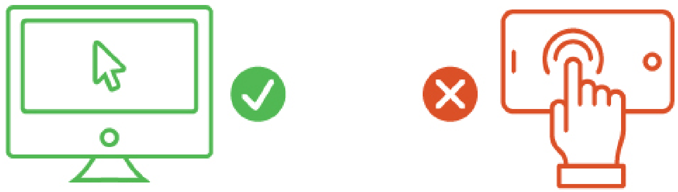


Figure 14: We included this image on the survey's first page to draw respondents' attention to the device requirements in order to participate in this study.

A total of 50 participants started the survey and provided their consent. Only 34, however, reached the survey's last page. We conducted data analyses using only information from these 34 participants and we deleted the other entries. A few hours after starting the data collection, we checked the first answers using `Python`. We discovered that the style preference had not been recorded for three participants due to a technical error in one condition (Group 2 with the order visualization A2 then B1). We promptly rectified this error and there was no other loss of data. Since these participants had completed the survey regarding the learning tasks and visualizations' ratings, we retained their data for the main inferential analyses and considered them as missing data for preference analysis (see Figure 15).

D.3.2 Attention checks

As pre-registered on `osf.io/ptsne` , we included two exclusion criteria to guarantee the quality of data before conducting any analysis. The first criterion consisted of passing at least one out of two attention checks in the rating scales; the second consisted of verifying that all answers to open-ended learning questions were sensical—that is, that participants showed they had indeed attempted to learn from the visualization, as required per the survey's instructions.

The first test consisted of attention check items included in the PREVis questionnaire, such as "For attention check purposes, please select **slightly agree** with this item". There was one attention check item per PREVis questionnaire. It was not technically possible to randomize their appearance; instead, the survey displayed them as part of different scales depending on the visualization (A1: among Understand items, A2: among Layout items, B1: among Data Feature items, and B2: among Data Read items). Each participant thus answered two attention checks: one required that they selected *slightly agree*, while the other required they chose the *slightly disagree*. We chose these labels from pre-test interviews conducted during the development of PREVis last year. At the time, the attention check items had more extreme choice requirements, and a few participants in that study mentioned they were not at ease with choosing from the extremes of a scale, even for attention check purposes. We are now careful to avoid extreme answer options in attention check items.

Six participants failed one attention check, but none of them failed both of them, which was our exclusion criteria. Therefore, we included all 34 participants for the second quality check, which consisted of verifying that all answers to open-ended learning tasks were sensical. A few answers did not entirely meet the instructions (for example, P2 provided statements instead of questions in one of the question tasks). All answers were, however, sensical. In addition, all but one were completely related to the topic at hand. The only exception was a comment from P5: ("Don't read a scientific paper when you're tired.") As a result, we did not exclude a single participant from the data analysis based on answer quality checks.

D.4 Conditions imbalance

In our study design, both the content and order of appearance of the two stimuli were randomized. However, due to participant dropout and the way our survey system assigned random conditions, the final number of complete submissions was uneven across conditions. Group 1 (A1 and B2) had $N = 15$ participants, while group B (A2 and B1) had $N = 19$ participants. Overall, 13 people saw Style 1 (compact) first and Style 2 (segmented) second, while 21 people saw Style 2 first and Style 1 second. This also shows we had a larger dropout rate on case where people saw Style 1 (compact) first Overall, 16 people saw content A first and content B second, while 18 people saw content B first and content A second.

Table 8: Participant distribution by randomized group, visualization order, and style order

Group	Order of stimuli content	Order of stimuli style	N
1 (A1 and B2)	A then B	1 then 2	**5**
	B then A	2 then 1	**10**
2 (A2 and B1)	A then B	2 then 1	**11**
	B then A	1 then 2	**8**

D.5 Style preferences

The segmented style was overall preferred; preferences were similarly distributed among self-reported levels of exposure to and familiarity with the topic.

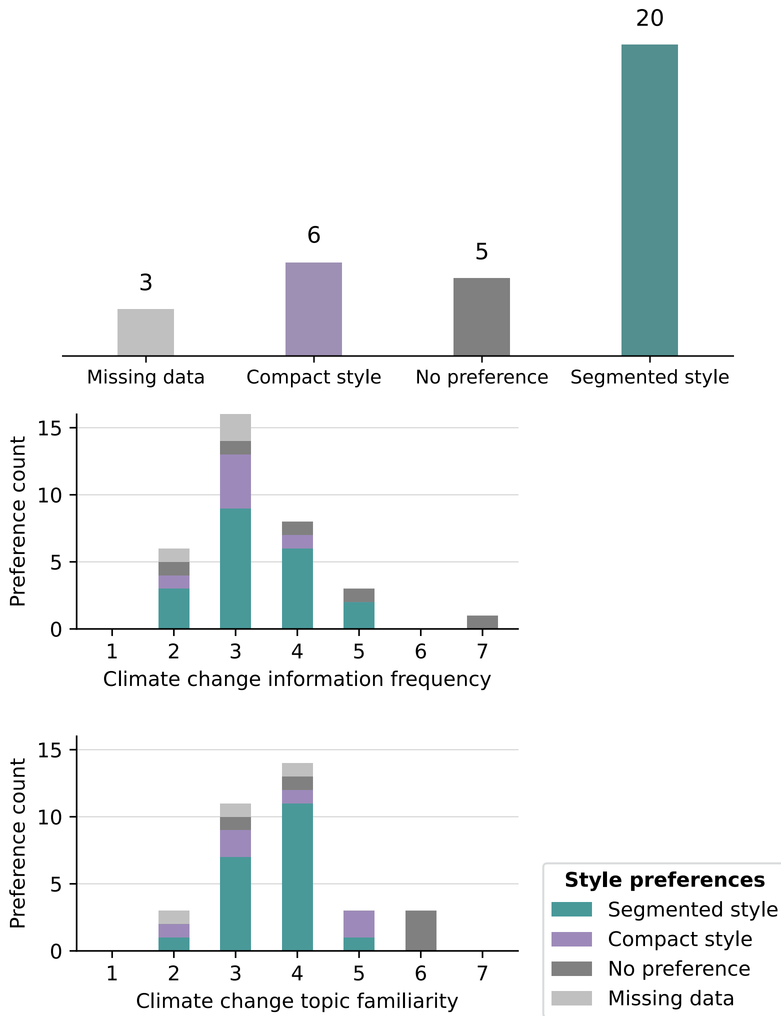


Figure 15: Top: style preferences distribution among all participants. Middle: style preferences by topic familiarity. Bottom: style preferences by exposure frequency.

E QUALITATIVE ANALYSIS OF PREFERENCE MOTIVATIONS

We conducted an inductive thematic analysis [14] of participants' comments regarding their style preferences. We adopted a realist perspective and carried out our analysis at the semantic level, focusing on participants' explicitly stated reasons in order to identify what they felt motivated their choices. We also focused solely on content that was related to our overarching question in this work, i. e., what participants experienced as easy or hard when retrieving information from the visualizations. As a result, we did not code the comment "NOTE: The image used for the compact presentation is different from the one shown in the survey." The first author conducted the analysis in multiple steps:

1. Fully read and re-read all comments. At that stage, the first author also read comments alongside the open-ended answers participants provided to tasks 1 and 2 for both stimuli, in order to contextualize participants' experiences.
2. Free coding across all comments, with a two-column structure: one for each style (compacted and segmented), as the goal was to identify what motivated participants' choices.
3. Establishing the codebook (see Figure 16 below) and harmonizing codes across all participants (see result on `osf.io/c73ys`).
4. Grouping codes into themes: this was an iterative process, merging and splitting themes, until finding coherent data groups. The first author used a simple table a visual representation, not a thematic map. We report the prevalence of themes among participants in Table 9.

code	codename	overall theme
clear	clear	clarity
compact is crowded	crowded	clarity
in segmented having to scroll back and forth is annoying	annoying	cost of interaction
scrolling costs time	scroll_time	cost of interaction
compact requires focus	focus_required	effort
segmenting makes it easier to read the data	easy_read	effort
easier to learn from a single view	easy_learn	effort
easy to understand	easy_understand	effort
had to read the compact version twice	read_twice	effort
less effort or mental strain	less_effort	effort
distracting	distracting	effort
segmented provides guidance	guidance	getting oriented
segmented with interaction allowed to recall where information was easily	recall	getting oriented
in compact it takes time to find key informations	key_lost	getting oriented
segmented provides key takeways up front	key_guidance	getting oriented
in the compact I have to find where to start	lost	getting oriented
focus on what is relevant	focus_guidance	getting oriented
compact can be overwhelming or discouraging	overwhelming	motivation
visually appealing	appealing	motivation
segmented gives information gradually	gradually	progression
overview first, then details	S_mantra	progression
segmented is like a story	story	progression
compact provides good overview of data - it's more practical all in one place	single_view	single view
text is distracting	text_distracting	text
compact has a lot of text	text_amount	text
more effort	more_effort	effort

Figure 16: Codes and themes from the participants' comments on their choice of preferred style.

Table 9: Theme prevalence among participants.

Theme	N
text	3
single view	4
motivation	5
cost of interaction	6
clarity	7
progression	12
getting oriented	19
effort	21

F Inferential data analysis

F.1 Repeated measures correlations

F.1.1 Correlation matrix

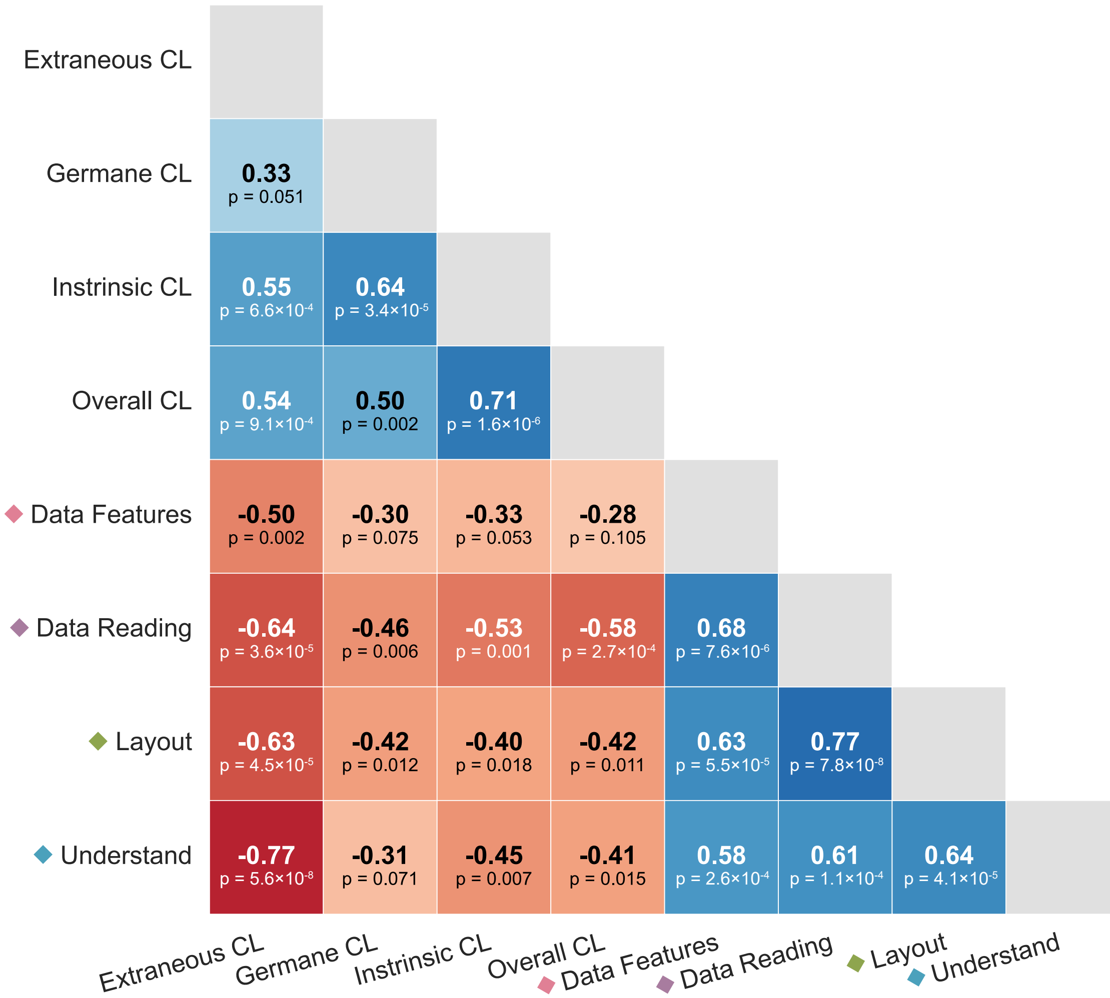


Figure 17: Repeated measures correlation matrix for subjective ratings of cognitive load (CL) and readability (PREVis) generated using the `rmcorr` package in R.

Figure 17 above represents the same data Figure 5, but we included all exact p values. For a more detail account, Table 10 on the next page provides repeated measures correlation coefficients and 0.95 confidence intervals, exact p values, and degrees of freedom for all pairs of variables.

As presented in the Results subsection 5.1, the correlation analysis reveals that perceived readability ratings negatively correlated with subjective ratings of extraneous cognitive load. Most negative correlations between extraneous cognitive load and PREVis scales [18] were strong, except for the **DataFeat** scale. The reliability indices for this particular scale were slightly weaker ($\alpha = 0.67$ and $\omega = 0.86$) than those of the other subscales (α ranged from 0.80 to 0.86, and ω ranged from 0.93 to 0.96, as reported in subsection 4.3). Lower reliability for the **DataFeat** scale could have affected the strength of its correlations with other measurements. This lack of reliability might be due to the study's small sample size combined with a lower number of items in the data features scale compared to other PREVis items (see Table 3). It is also possible that these items caused confusion among respondents as their wording is more complex than that of other PREVis items.

Further exploration of this matrix shows that the **DataRead** and **Layout** ratings also moderately and negatively correlated with germane, intrinsic, and overall cognitive load ratings ($r \leq$ -.40, $p < 0.05$). In addition, PREVis **Understand** ratings moderately correlated with intrinsic and overall cognitive load ratings ($r \leq -.41$, $p < 0.05$). There were no other significant correlations between readability and cognitive load ratings.

Finally, PREVis ratings moderately to strongly correlated amongst each other ($r \geq .58$, $p < 0.001$), and similarly for cognitive load ratings among each other ($r \geq .50$, $p < 0.01$), to the exception of germane and extraneous loads for which the correlation was weak and non-significant ($r = .33$, $p > 0.05$).

Table 10: Detailed repeated measures correlation values: degree of freedom for each pair, repeated measures correlation coefficient, CI boundaries (at the 0.95 level), and exact p values.

Measure 1	**Measure 2**	**df**	r_{rm}	**Lower CI**	**Upper CI**	p **value**
Overall CL	Intrinsic CL	33	0.71	0.50	0.84	1.6×10^{-6}
Overall CL	Extraneous CL	33	0.54	0.25	0.74	0.00091
Overall CL	Germane CL	33	0.50	0.20	0.71	0.0022
Overall CL	◆ Understand	33	-0.41	-0.65	-0.08	0.015
Overall CL	◆ Layout	33	-0.42	-0.66	-0.10	0.011
Overall CL	◆ Data features	33	-0.28	-0.56	0.06	0.11
Overall CL	◆ Data read	33	-0.58	-0.76	-0.30	0.00027
Intrinsic CL	Extraneous CL	33	0.55	0.26	0.75	0.00066
Intrinsic CL	Germane CL	33	0.64	0.39	0.80	3.4×10^{-5}
Intrinsic CL	◆ Understand	33	-0.45	-0.68	-0.14	0.0067
Intrinsic CL	◆ Layout	33	-0.40	-0.65	-0.08	0.018
Intrinsic CL	◆ Data features	33	-0.33	-0.60	0.00	0.053
Intrinsic CL	◆ Data read	33	-0.53	-0.73	-0.24	0.0011
Extraneous CL	Germane CL	33	0.33	-0.00	0.60	0.051
Extraneous CL	◆ Understand	33	-0.77	-0.88	-0.59	5.6×10^{-8}
Extraneous CL	◆ Layout	33	-0.63	-0.80	-0.38	4.5×10^{-5}
Extraneous CL	◆ Data features	33	-0.50	-0.71	-0.20	0.0023
Extraneous CL	◆ Data read	33	-0.64	-0.80	-0.39	3.6×10^{-5}
Germane CL	◆ Understand	33	-0.31	-0.58	0.03	0.071
Germane CL	◆ Layout	33	-0.42	-0.66	-0.10	0.012
Germane CL	◆ Data features	33	-0.30	-0.58	0.03	0.075
Germane CL	◆ Data read	33	-0.46	-0.69	-0.15	0.0059
◆ Understand	◆ Layout	33	0.64	0.38	0.80	4.1×10^{-5}
◆ Understand	◆ Data features	33	0.58	0.31	0.77	0.00026
◆ Understand	◆ Data read	33	0.61	0.35	0.78	0.00011
◆ Layout	◆ Data features	33	0.63	0.37	0.79	5.5×10^{-5}
◆ Layout	◆ Data read	33	0.77	0.58	0.88	7.8×10^{-8}
◆ Data features	◆ Data read	33	0.68	0.45	0.83	7.6×10^{-6}

F.2 Paired-samples ttests

F.2.1 Effect of presentation style on cognitive load ratings

Table 11 provides descriptive values for cognitive load measurements across styles. As reported in the Results section 5.1, we conducted a series of paired-samples *t*-tests to assess the effect of presentation style on cognitive load ratings. The results are detailed in Table 12 below. There was a significant difference in extraneous cognitive load between compact (M = 4.44, SD = 1.85) and segmented (M = 3.56, SD = 1.44) styles, $t(33) = -2.520$, $p < .01$. There were no significant differences for overall, intrinsic, and cognitive load measurements.

Table 11: Descriptive statistics for cognitive load ratings between compact presentation style (style 1) and segmented presentation style (style 2).

	N	Mean	Median	SD	SE
Style1 CL Extraneous	34	4.44	4.00	1.85	0.32
Style2 CL Extraneous	34	3.56	3.00	1.44	0.25
Style1 CL Overall	34	5.59	6.00	1.50	0.26
Style2 CL Overall	34	5.21	6.00	1.39	0.24
Style1 CL Intrinsic	34	5.09	5.50	1.54	0.26
Style2 CL Intrinsic	34	4.74	5.00	1.44	0.25
Style1 CL Germane	34	5.68	6.00	1.41	0.24
Style2 CL Germane	34	5.41	6.00	1.35	0.23

Table 12: Paired sample *t*-tests for cognitive load ratings between style 1 (compact presentation) and style 2 (segmented presentation).

								95% Confidence Interval	
			statistic	df	p		Effect Size	Lower	Upper
Style1 CL Extraneous	Style2 CL Extraneous	Student's t	2.520	33.0	0.008	Cohen's d	0.432	0.0774	0.781
Style1 CL Overall	Style 2CL Overall	Student's t	1.601	33.0	0.060	Cohen's d	0.274	-0.0700	0.615
Style1 CL Intrinsic	Style2 CL Intrinsic	Student's t	1.015	33.0	0.159	Cohen's d	0.174	-0.1660	0.511
Style1 CL Germane	Style2 CL Germane	Student's t	0.828	33.0	0.207	Cohen's d	0.142	-0.1969	0.479

Note. H_a $\mu_{\text{Measure 1 - Measure 2}} > 0$

Having found a significant difference in extraneous cognitive load between compact and segmented styles, we conducted a post hoc repeated measures analysis of variance to assess whether the visualizations' content (`rand_group`) or their order of appearance (`vis_order`) significantly affected participants' ratings of extraneous cognitive load, or if these factors interacted with the presentation style. We used the random group variable to test the effect of content because the visualizations' content for a given style was different depending on the random group assigned to participants (see Figure 4). Results in Tabs. 13 and 14 showed that only the primary within-participant effect of style was significant, albeit very small ($F(1,30) = 6.11, p < .05, \eta_p^2 = 0.17$). There were no other significant effects or interactions between factors: for a given style, the visualizations' content and appearance order did not affect learners' extraneous cognitive load.

Table 13: Repeated Measures ANOVA: Within-subjects effects of style, order of display (`vis_order`), and visualization content (`rand_group`) on extraneous cognitive load.

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_p
Style	13.53	1	13.53	6.11	0.019	0.17
Style ✻ vis_order	0.21	1	0.21	0.09	0.761	0.00
Style ✻ rand_group	0.28	1	0.28	0.13	0.725	0.00
Style ✻ vis_order ✻ rand_group	1.55	1	1.55	0.70	0.410	0.02
Residual	66.47	30	2.22			

Note. Type 3 Sums of Squares

Table 14: Repeated Measures ANOVA: Between subjects effects of order of display (`vis_order`) and visualization content (`rand_group`) on extraneous cognitive load.

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_p
vis_order	0.04	1	0.04	0.01	0.912	0.00
rand_group	5.07	1	5.07	1.43	0.241	0.05
vis_order ✻ rand_group	0.94	1	0.94	0.27	0.610	0.01
Residual	106.22	30	3.54			

Note. Type 3 Sums of Squares

F.2.2 Effect of presentation style on readability ratings

Table 15 provides descriptive values for readability measurements across styles. Having found significant negative correlations between extraneous cognitive load and readability measures as well as a significant effect of style on extraneous cognitive load, we conducted a series of exploratory paired-samples t-tests to assess the effect of presentation style on readability, as shown in Tab. 16. There were significant differences between compact and segmented styles for the ◆ **Understand**, ◆ **Layout**, and ◆ **DataRead** readability scales but not for the ◆ **DataFeat** scale. The reliability indices for this particular subscale were slightly weaker ($\alpha = 0.67$ and $\omega = 0.86$) than those of the other subscales (α ranged from 0.80 to 0.86, and ω ranged from 0.93 to 0.96, as reported in subsection 4.3). Lower reliability for the ◆ **DataFeat** scale could have affected the strength of its correlations with other measurements. This lack of reliability might be due to the study's small sample size combined with a lower number of items in the data features scale compared to other PREVis items (see Table 3). It is also possible that these items caused confusion among respondents as their wording is more complex than that of other PREVis items.

Table 15: Descriptive statistics for perceived readability ratings between Compact stimuli (Style 1) and Segmented stimuli (Style 2).

	N	Mean	Median	SD	SE
Style **1** Understand ◆	34	4.86	5.00	1.51	0.26
Style **2** Understand ◆	34	5.56	5.67	1.20	0.21
Style **1** Layout ◆	34	4.04	3.83	1.68	0.29
Style **2** Layout ◆	34	5.21	6.00	1.56	0.27
Style **1** DataRead ◆	34	4.57	5.00	1.62	0.28
Style **2** DataRead ◆	34	5.20	5.50	1.31	0.22
Style **1** DataFeat ◆	34	4.81	5.00	1.66	0.28
Style **2** DataFeat ◆	34	5.25	5.50	1.49	0.26

Table 16: Paired sample t-tests for perceived readability ratings between Compact stimuli (Style 1) and Segmented stimuli (Style 2).

								95% Confidence Interval	
			statistic	df	p		Effect Size	Lower	Upper
Style **1** Understand ◆	Style **2** Understand ◆	Student's t	-2.69	33.0	0.006	Cohen's d	-0.461	-0.812	-0.1041
Style **1** Layout ◆	Style **2** Layout ◆	Student's t	-3.64	33.0	< .001	Cohen's d	-0.625	-0.989	-0.2524
Style **1** DataRead ◆	Style **2** DataRead ◆	Student's t	-2.19	33.0	0.018	Cohen's d	-0.376	-0.721	-0.0252
Style **1** DataFeat ◆	Style **2** DataFeat ◆	Student's t	-1.58	33.0	0.062	Cohen's d	-0.270	-0.611	0.0740

Note. H_a $\mu_{\text{Measure 1 - Measure 2}} < 0$

G Time spent

Participants tended to spend less time on tasks in the second visualization they saw, regardless of the style or content. However, multiple issues limit the insights we can derive from our time records:

- Some participants did not complete the survey in one go. To compute the average time spent by participant in the study, we removed one outlier whose recorded overall time was 545 min.

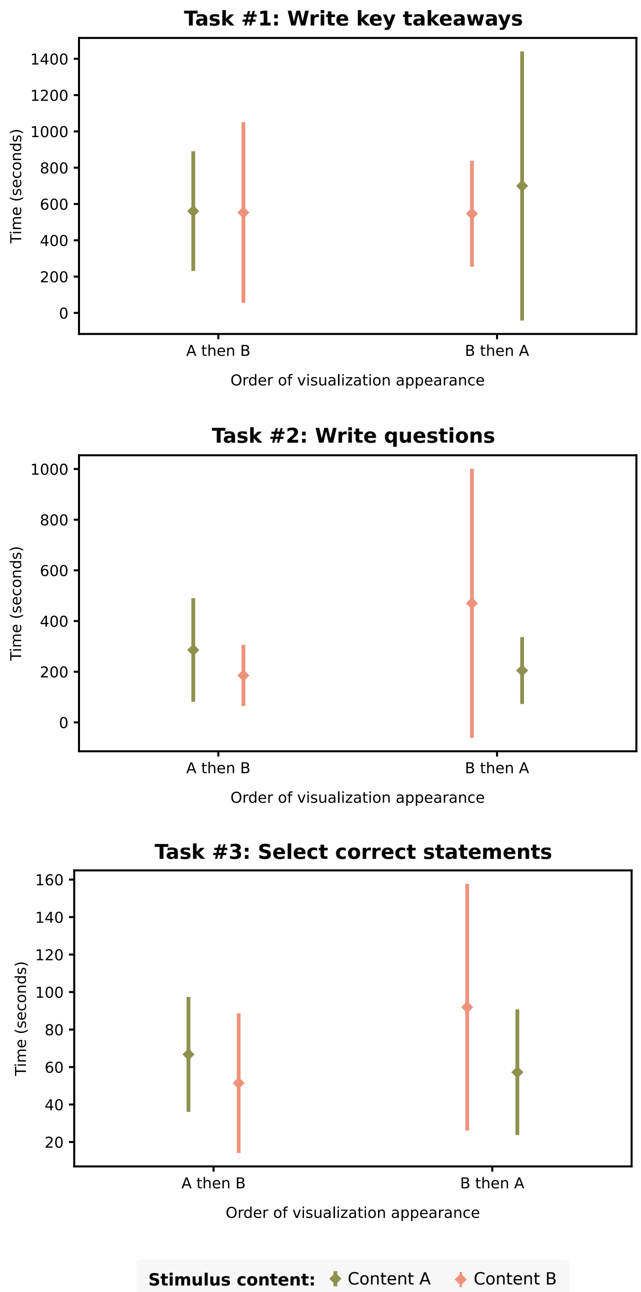


Figure 18: Point estimate with 95% CI of time spent on tasks per content of presentation (Content A = heat-humidity risks to human health, Content B = risk of species losses), separated by order of appearance.

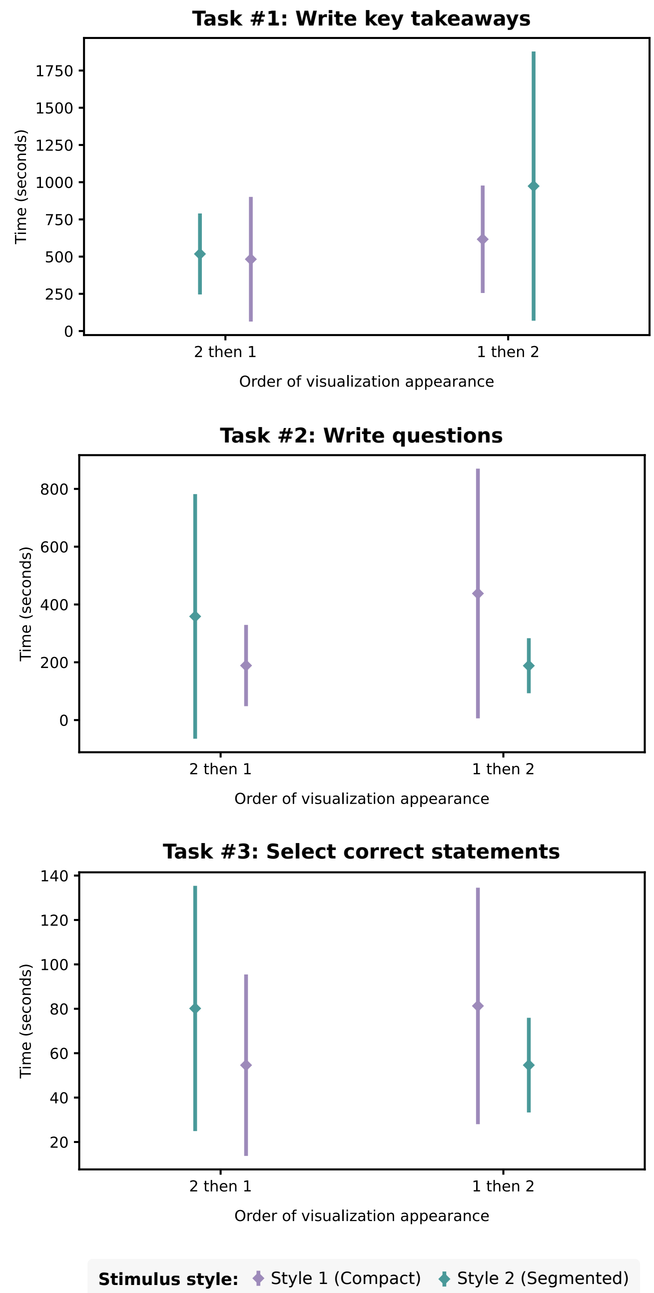


Figure 19: Point estimate with 95% CI of time spent on tasks per style of presentation (Style 1 = Compact, Style 2 = Segmented), separated by order of appearance.

H TASK PERFORMANCE BY PRESENTATION STYLE

As our focus was on cognitive load measurements and not trask performance, we did not conduct a deep analysis of the tasks' outcomes (see Sec. 4.2 for a description of learning tasks). For open-ended tasks #1 and #2, task performance was only evaluated by marking statements and questions that were *relevant* to the stimulus, that is, directly related to the content presented in the visualization and accompanying text.

We report below the task performance across different levels of self-reported familiarity with climate change (labeled *CCFamiliarity* in the figures) and frequency of exposure to climate change visualizations (labeled *CCFrequency* in the figures). However, given the small sample size, limited control over experimental conditions, and the absence of clear performance differences, it is difficult to draw any meaningful conclusions about the impact of presentation style on task outcomes.

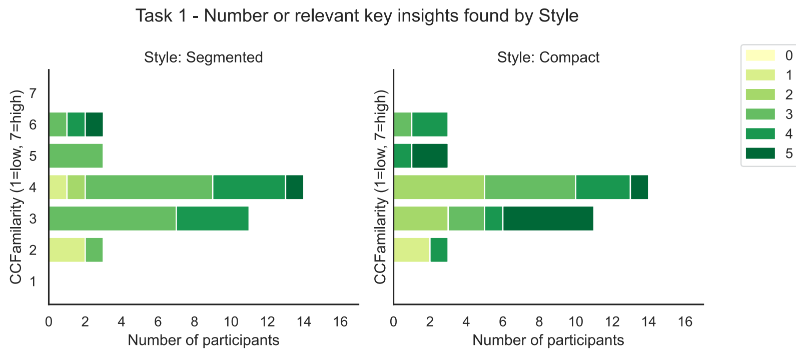


Figure 20: Task performance: Number or relevant key insights found performance by CCFamilarity by Style

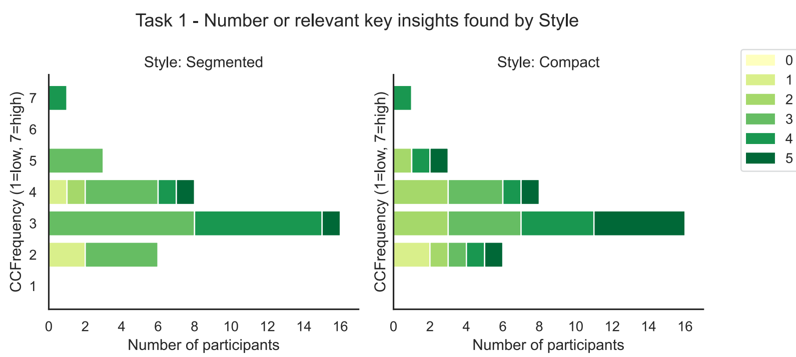


Figure 21: Task performance: Number or relevant key insights found performance by CCFrequency by Style

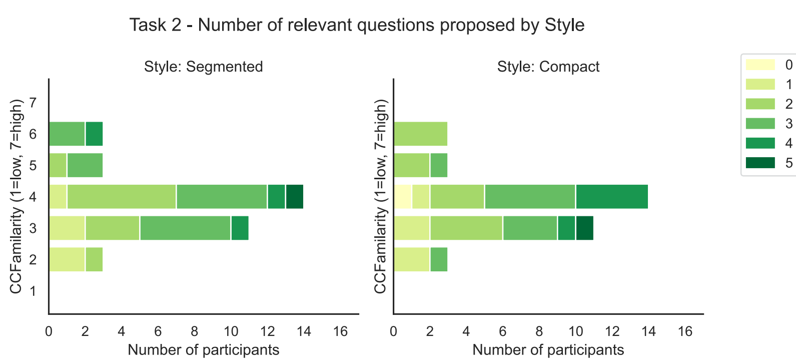


Figure 22: Task performance: Number of relevant questions proposed performance by CCFamilarity by Style

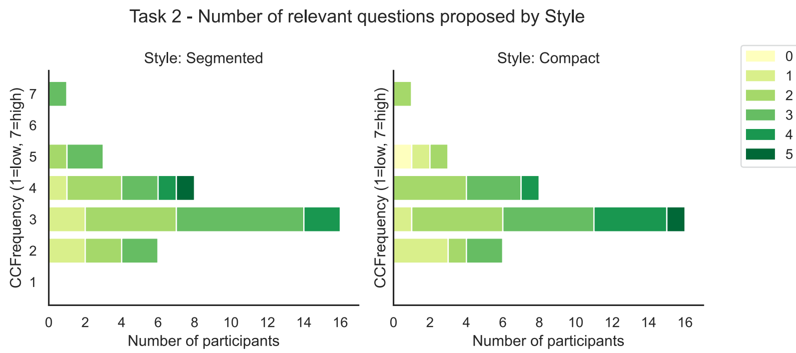


Figure 23: Task performance: Number of relevant questions proposed performance by CCFrequency by Style

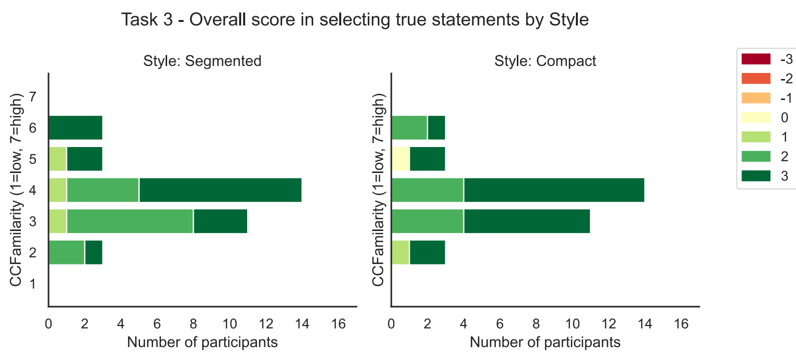


Figure 24: Task performance: Overall score in selecting true statements performance by CCFamilarity by Style

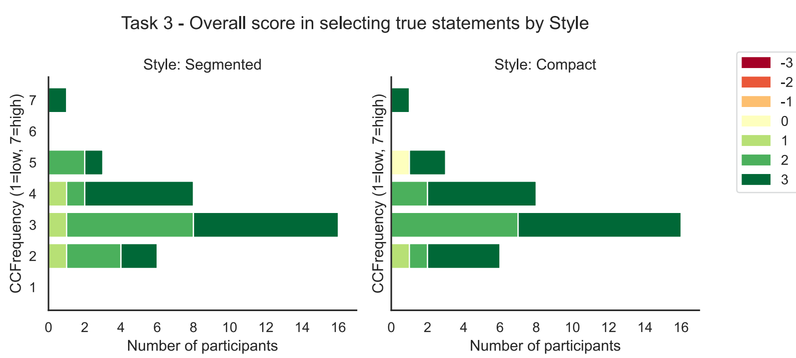


Figure 25: Task performance: Overall score in selecting true statements performance by CCFrequency by Style

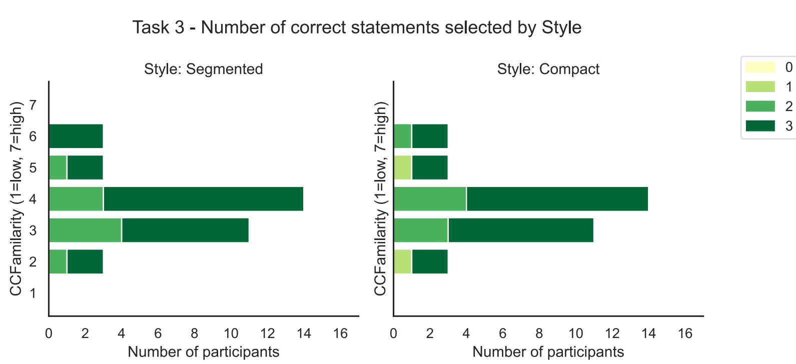


Figure 26: Task performance: Number of correct statements selected performance by CCFamilarity by Style

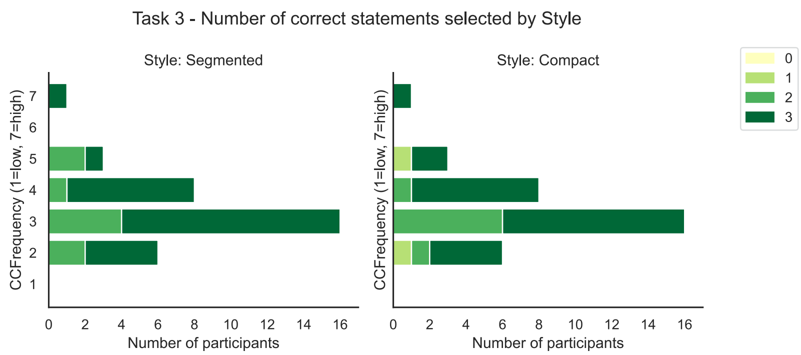


Figure 27: Task performance: Number of correct statements selected performance by CCFrequency by Style

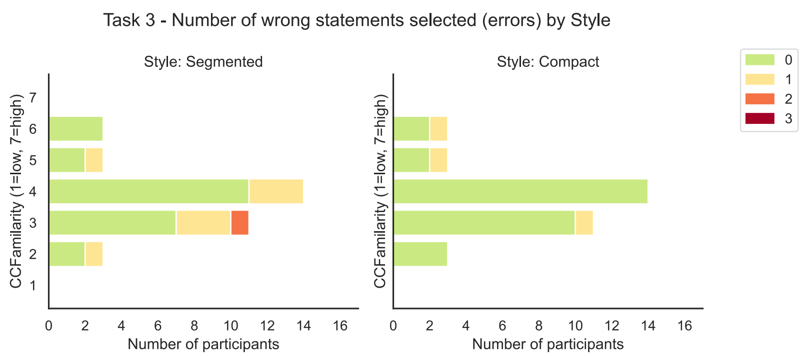


Figure 28: Task performance: Number of wrong statements selected (errors) performance by CCFamilarity by Style

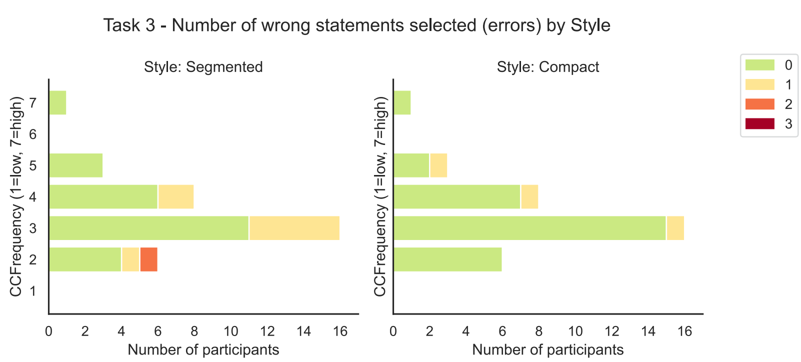


Figure 29: Task performance: Number of wrong statements selected (errors) performance by CCFrequency by Style