



HAL
open science

A cognitive load approach to designing and evaluating data visualizations

Anne-Flore Cabouat

► **To cite this version:**

Anne-Flore Cabouat. A cognitive load approach to designing and evaluating data visualizations. Cognitive science. 2024. hal-04885430

HAL Id: hal-04885430

<https://inria.hal.science/hal-04885430v1>

Submitted on 14 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

A cognitive load approach to designing and evaluating data visualizations

Student: Anne-Flore Cabouat

Université Paris Cité

EURIP graduate school

Learning Planet Institute

Master's Thesis

Supervisors: Lorenzo Ciccione, Petra Isenberg, Tobias Isenberg

June 10th, 2024

Abstract

Visual representations of data are increasingly prevalent, but we lack a detailed theoretical framework to explain what factors make easy or difficult to read and understand; nor do we know how such factors can impact data visualizations' efficiency as learning material. In this Master's Thesis, I start to address this gap by exploring the validity and applicability of Cognitive Load Theory, an educational and cognitive science theoretical framework, for designing and evaluating data visualizations. To that end, I conducted an online randomized study in which each participant (N=34) performed learning tasks on two different data visualizations from the 6th IPCC Summary for Policy Makers. One was presented in three successive parts, according to the "segmenting" principle in 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. I found significant negative correlations between readability and extraneous cognitive load, indicating that learners felt they needed to put less effort into learning from a visualization when they found it more readable. In addition to a qualitative analysis of learners' preferences, in this work I also contribute an interdisciplinary perspective on visualization cognitive processing and a discussion of implications for evaluating readability in visualization studies.

Highlights:

- Visualization readability was strongly and negatively correlated with extraneous cognitive load in learners;
- Segmenting data visualizations did not significantly reduce overall mental effort in learners;
- A majority of participants preferred the segmented style as they felt guided in reading and understanding the content, but the need to scroll between parts of the visualization could also feel cumbersome at times.

Keywords: Data visualization, Cognitive load, Readability, Multimedia learning

In this document, I use the following abbreviations: **CLT** for Cognitive Load Theory, and **CTML** for Cognitive Theory of Multimedia Learning. Please also note that figures and tables from the main part of this work use simple number references, while figures and tables from the Appendices use the Appendix letter with a number. For example, [Figure C2](#) is the second figure in Appendix C, while [Figure 2](#) is the second figure in the main paper.

A cognitive load approach to designing and evaluating data visualizations

Introduction

Visual representations of data are increasingly used for learning, communicating, or daily decision-making (Munzner, 2014). Data visualizations facilitate understanding of their underlying datasets by substituting complex cognitive processes for simple visual pattern identification, thus serving as external cognition tools (Bertin, 1967 (2011); Card *et al.* 1999). When used in learning contexts, their form influences the structure of mental models that learners build about the represented phenomenon (Schnotz & Kürschner, 2008). As data visualizations became ubiquitous, researchers from various domains such as computer science (Borkin *et al.* 2013; Burns *et al.* 2022), psychology (Fansher *et al.* 2022; Shah & Freedman, 2011), media and communication (Sánchez-Holgado *et al.* 2023), health decision making (Galesic & Garcia-Retamero, 2011), education (Curcio, 1987; Shah & Hoeffner, 2002), and neuroscience (Ciccione, 2022) began to study how people understand and learn from data visualizations. These research communities rely on diverse and complementary expertise and methodologies, yet they too rarely interact (Franconeri *et al.* 2021). In this work, I aim to contribute to bridging this gap by using a cognitive load and instructional design perspective to create and analyze data visualizations for learning situations. In the following subsections, I first discuss relevant literature on learning from data visual representations and then introduce important concepts from instructional design literature.

Learning from data visualizations

Scholars have studied how people use data visualization to acquire new knowledge in two main research domains: *data visualization* in computer science and *graph comprehension* in educational psychology. In this subsection, I briefly review relevant work from both domains.

Data visualization domain

In the computer science field, data (or information) visualization is defined as “the use of computer-supported, interactive, visual representations of abstract data” (Card *et al.* 1999). Early research mostly focused on visual analytics with domain expert users (Wong *et al.* 2018), that

is, people who are already highly knowledgeable about what the represented data refers to. As a result, the use of data visualization as educational material was not a prevalent topic in the field, and few publications explicitly drew from the learning science literature. Zhu (2007) suggested measuring peoples' learning curve as a way to assess data visualization efficiency; more recently, Burns *et al.* (2020) proposed to adapt Bloom's (1956) taxonomy of educational goals to identify and assess different levels of understanding in data visualization viewers, and a research team (Adar & Lee, 2021; Lee-Robbins *et al.* 2022) proposed the use of learning objectives as a specification framework for designing and evaluating visualizations.

Beyond these rare publications that explicitly draw from educational sciences, a few studies focus on visualization understanding, recall, and recognition. Lee *et al.* (2015), for example, investigated how people make sense of unfamiliar visualizations; Borgo *et al.* (2012) examined how to design visualizations to support readers' understanding and memorization; and Borkin *et al.* (2013, 2016) studied memorability of data visualizations. Although the aforementioned studies were not conducted in educational settings, their measured outcomes are comparable to learning goals as Jarodzka (2021) defined them in the context of multimedia learning: remembering (i. e., retention) and understanding (i. e., transfer). Educational theories are thus likely to provide useful insights to the visualization field at large, which motivated the research I present next.

Moreover, cognitive theories from the instructional design literature can offer researchers theoretical perspectives on improving data visualization design. Indeed, data visualization research as a field is mostly focused on gathering empirical results but lacks comprehensive theories to explain and predict how individuals process data visual representations (Fox & Hollan, 2023). In particular, although the community gathered a vast amount of factual knowledge about graphical perception (Franconeri *et al.* 2021; Quadri & Rosen, 2022; Zeng & Battle, 2023), there is no theoretical framework available to explain such observations. For instance, empirical studies (Pandey *et al.* 2015; Correll *et al.* 2020) showed that visualization readers tend to overestimate differences between bars when a chart relies on truncated axes (see Figure 1). Still lacking are models that

explain the cognitive or psychological mechanisms at play in such empirical findings. Psychology researchers studying *graph comprehension* have attempted to derive such explanatory models, which I briefly describe next.

Graph comprehension domain

The *graph comprehension* field in psychology research involves studying how individuals understand visual representations of data, which are commonly referred to as “graphs” or “charts.” Seminal work in this field defined these representations as visual objects where data values are mapped to visual dimensions, such as length, position, color, or shape (Bertin, 1967 (2011); Pinker, 1990), and provided early models to explain how individuals process such visual information (Kosslyn, 1989; Pinker, 1990). Researchers in education also explored how such representations can assist learners in understanding quantitative information (DeSanctis & Jarvenpaa, 1985). They refined cognitive models to explain the processes involved in understanding visual representations of data (Carpenter & Shah, 1998) and found that the factors influencing graph comprehension were similar to factors that influence text comprehension, namely: visual characteristics of the object (graph or text), viewers’ *domain knowledge* (i. e., knowledge of the topic), literacy (or graphical literacy), and reasoning skills. As a result, Freedman and Shah (2002) proposed to use a model of text comprehension to explain graph comprehension. Hegarty (2011) summarized Pinker (1990) and Freedman and Shah (2002) in the model presented in Figure 2.

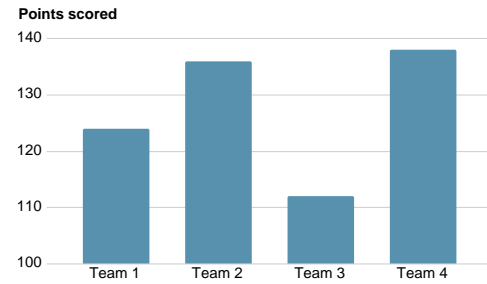


Figure 1. Example of a misleading bar chart, where using a truncated y-axis visually amplifies differences in scores.

Cognitive model of visualization comprehension

→ Cognitive process □ Internal representation

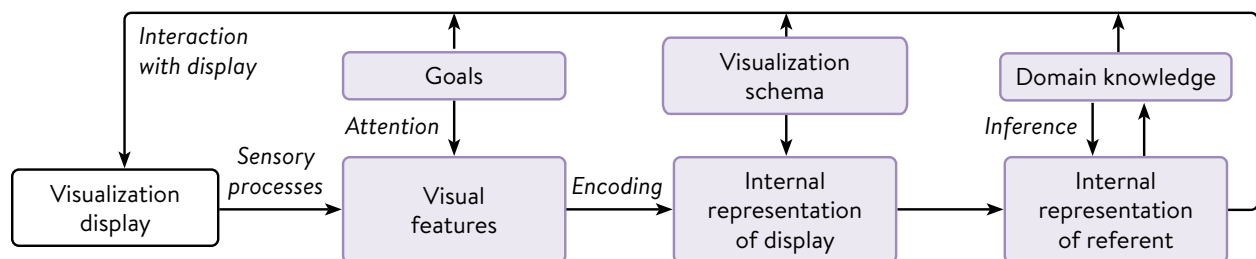


Figure 2. Internal representations and processes involved in understanding a visualization (Hegarty, 2011).

Researchers from the domain, however, did not investigate the application of instructional design theories to data visualizations. In this work, I intend to establish clear connections between data visualization design and important theories and concepts relevant to designing learning materials, which I will introduce next.

Instructional design frameworks

Cognitive Load Theory (CLT)

Although Sweller (1988) first developed CLT as a framework for instructional design in problem-solving tasks, it was extended to encompass a wider range of learning tasks (Sweller *et al.* 2011). CLT is organized around the concept of *cognitive load*, which represents the workload a learning task imposes on the learner's cognitive system (Paas *et al.* 2003). There are three main cognitive load components in CLT: *intrinsic load*, *extraneous load*, and *germane load*—also known as *relevant information load* (Paas & Sweller, 2021). ***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 cannot be altered other than by changing the learning objective or by increasing the learner's prior knowledge and skills. ***Extraneous load***, in contrast, is solely imposed by the design of learning material and activities. Once intrinsic and extraneous loads have been withdrawn from a learner's cognitive resources pool, the remaining resources can be allocated to knowledge acquisition—i. e., updating mental representations of the world with new, relevant information. This use of cognitive resources is called ***germane load***. CLT also relies on the assumption that a learner's individual cognitive resources pool cannot be increased. Therefore, learning cannot occur effectively if the sum of intrinsic and extraneous cognitive loads exceeds the learner's cognitive capacity limits. CLT thus 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 learning rather than to processing extraneous information.

Cognitive Theory of Multimedia Learning (CTML)

Sweller's CLT was foundational to inspire the work of Mayer on a Cognitive Theory of Multimedia Learning (Camp *et al.* 2021; Mayer, 2001a)—which, in turn, influenced the evolution

of the CLT framework (Sweller *et al.* 2011). Extended research has been conducted over the past 20 years to integrate, refine, and extend CTL for the specific context of multimedia learning in CTML (Mayer & Fiorella, 2021a). This work resulted in an evolving set of design principles, which provide an elaborate framework dedicated to multimedia instructional design. In this context, Mayer (2021) broadly defines multimedia instructional material as “a communication containing words and pictures intended to foster learning.” The main design principle in CTML is the multimedia principle (Mayer, 2001b), which states that people learn better from words and pictures than from words alone. Other principles in CTML define conditions to optimize multimedia learning outcomes (Mayer & Fiorella, 2021a). Those guidelines aim to reduce extraneous load, manage essential learning (which relates to germane load), engage the learner, and integrate social and affective components in instructional design.

CTL and CTML applied to data visualizations

The concept of cognitive load is not new to visualization researchers, although it is not often discussed. For example, Anderson *et al.* (2011) and Yoghourdjian *et al.* (2021) used electroencephalogram (EEG) measurements of cognitive load with different visualizations. While Anderson *et al.* (2011) focused on extraneous cognitive load after controlling for germane load and intrinsic load in their experimental design, Yoghourdjian *et al.* (2021) did not attempt to obtain distinct measures of extraneous, intrinsic, and germane load. As CLT predicts the effects of instructional design manipulations on specific types of cognitive load, I included distinct measurements for overall, intrinsic, extraneous, and germane cognitive loads in my study design.

If cognitive load measurements are rare in data visualization user studies, studies involving the application of design principles from CLT or CTML are even rarer. To the best of my knowledge, only Kopp *et al.* (2018) investigated this matter, but the authors focused on task performance measures instead of cognitive load or learning outcomes. Surprisingly, such studies are also rare in the education literature. While research in multimedia learning often uses visual representations such as diagrams to aid learning (McCrudden & Van Meter, 2021), data visualizations have not been a major focus in this field. Even when studying learning activities with data visualizations,

educational science researchers did not collect cognitive load measurements. A study by Mautone and Mayer (2007), for example, 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 did not assess whether such application of CTML design principles to graphs had an impact on learners' cognitive load. To fill that gap, my controlled study focuses on assessing how applying a CTML design principle to data visualizations can affect viewers' cognitive load.

If we think of data visualization as a learning problem, we can use CTML principles to improve the design of data visual representations. We can also evaluate visualization activities as learning experiences using cognitive load measurements. Although visualization researchers Huang *et al.* (2009) and Adar and Lee (2021) proposed somewhat similar ideas, they did not put forward a theoretical analysis of how various visualization characteristics might affect different types of cognitive load in viewers. To bridge this gap, I propose such an analysis below and examine how we can connect concepts from the CLT and CTML framework to data visualization literature.

Factors of cognitive load in data visualization

With this work, my goal is to determine how relevant the CTML and CLT frameworks are for data visualization design and quality assessment. CLT is, in essence, relevant to the presentation of any complex information that requires a reduction in processing workload (Sweller, 2020). Early work on visualization complexity showed that visual representations of data can deliver overwhelming amounts of information (Brath, 1997). Some of this complexity can be attributed to the nature of underlying data, some to the appropriateness and efficiency of the visual design, and some to the viewer's knowledge and skill in reading data visualizations (Isenberg, 2021).

In CLT and CTML, *element interactivity* describes the amount of novel elements of information that a learner needs to process simultaneously. The level of relevant element interactivity is the primary factor that determines *intrinsic load* in learning tasks. As multiple elements can be combined into a single element during learning (e.g., learning how to read allows the transformation of the individual letters "c", "a", and "t" into a single word "cat"), intrinsic load is mainly influenced by the learner's prior knowledge (Paas & Sweller, 2021). 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 (Larkin & Simon, 1987)—but a reader's domain knowledge could have an effect on 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*. If all represented data dimensions are relevant to the learning goal, a higher number of data dimensions would thus directly lead to a higher intrinsic cognitive load in data visualization.

Such data dimensions, 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 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 (Ghoniem *et al.* 2005; Thudt *et al.* 2016). 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). Furthermore, as Quadri and Rosen (2022) emphasized, visualization effectiveness is task-dependant. This means that the adequacy of the representation for supporting visualization *tasks* (Brehmer & Munzner, 2013) relevant to the learning goals will also impact readability—and extraneous load. The characteristics of learners also play a crucial role in readability (Fox & Hollan, 2023). Learners with low data visualization literacy (Boy *et al.* 2014; Lee *et al.* 2017) may struggle to interpret data visual representations, thus experiencing lower readability (Cabouat *et al.* 2024a)—and higher extraneous cognitive load. In summary, we can generally expect a data visualization to generate less extraneous load when it is more readable. My study design includes measurements of perceived readability to help bridge the gap between the different streams of research I presented in this section. A strong correlation between participants' ratings of extraneous cognitive load and perceived readability would indicate that CLT is a relevant framework not only for designing but also 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. Conversely, well-designed features can provide learners with greater control (Scheiter & Gerjets, 2007) over the pace of information display and on-demand access to more detailed information. This can help learners manage the allocation of cognitive resources to *germane load* more effectively (Mayer & Fiorella, 2021b). More specifically, studies have shown that *segmenting* the presentation of information can reduce the overall cognitive load (Rey *et al.* 2019). To the best of my knowledge, however, the effects of the segmenting principle have not been studied yet in the context of data visualization. In this work, I will address this gap by comparing cognitive load measurements between a control and a segmented data visualization.

Research question and hypothesis

To summarise, I found strong indications in the literature that Cognitive Load Theory (CTL) should provide a relevant framework to assess visualization efficiency. Accordingly, the derived Cognitive Theory of Multimedia Learning (CTML) should also provide relevant design principles for improving data visualizations. In this work, I want to address the following research question:

To what extent do CTL and CTML apply to information retrieval from data visualizations?

I 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 instructional design principles from CTML to data visualizations decreases the overall cognitive load in learners.

Method

I conducted a pre-registered (osf.io/ptsne) cross-sectional experimental study with a within-subject design. My goal was to assess the effects of applying an instructional design principle to the presentation of data visualizations on learners' cognitive load and to assess the correlation between perceived visualization readability and learners' subjective extraneous

cognitive load. I collected subjective ratings of readability and cognitive load from learners on two different data visualization stimuli: a control stimulus, and a stimulus presented according to the segmenting principle from the Cognitive Theory of Multimedia Learning (CTML). To gain a more comprehensive understanding of participants' learning experiences with the visualizations, I also collected their feedback about which visualization they preferred and why.

Participants

My advisors and I 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 (mobile devices were not supported). We sent invitations starting from April 24th, informing participants that the study would run until May 4th. As pre-registered, I closed the survey on that date because I had reached my 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. As a result, these participants were not included in the qualitative analysis, but their other answers in the survey were still used for quantitative analysis. 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. I provide a complete table of demographics data in [Table D1](#). Participant's 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*; see [Table D3](#) for a detailed list of answer options and frequencies). 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 [Table D2](#) for a detailed list of answer options and frequencies).

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. I derived these tasks from the work of Quadri *et al.* (2024), who found that different tasks elicit different response patterns in people, in line with existing survey design recommendations (Krosnick, 2017).

Data visualizations. Each participant completed learning tasks on two different data visualizations (see [Figure 3](#)). I chose to use existing visualizations in order to recreate conditions as close as possible to real-world scenarios. Specifically, I chose visualizations related to climate change to better engage the participants in their learning tasks compared to artificial content. Such visualizations are also likely to elicit high cognitive load in learners: existing studies showed that climate visualizations can be hard to understand for non-experts (McMahon *et al.* 2015) and that designing climate change data visualizations presents challenges in overcoming visual complexity for authors from the Intergovernmental Panel on Climate Change (IPCC) (Harold *et al.* 2020). I chose to use *density maps* because the Visualization Literacy Assessment Test items have low difficulty and discriminatory power (Lee *et al.* 2017) for this type of representation, and I wanted to control the effect of participants' visualization literacy on extraneous cognitive load. I used two visualizations from Figure SPM.3 in the IPCC 6th Assessment Synthesis Report Summary for Policy Makers (Calvin *et al.* 2023). Figure SPM.3 (see [online](#) version) presents three topics with multiple density maps each. I chose the first two topics because they had very similar layouts, showed four maps each, and used the same colors, thus minimizing the possible effects of graphic design on extraneous load. Both images showed maps with projections across different scenarios of global warming for heat-humidity risks to human health (**content A**: see [Figure A1](#)) and species loss risk (**content B**: see [Figure A2](#)). 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, I created additional text information, which I based on the synthesis report (Calvin *et al.* 2023), the underlying studies they cited, and IPCC’s full reports.

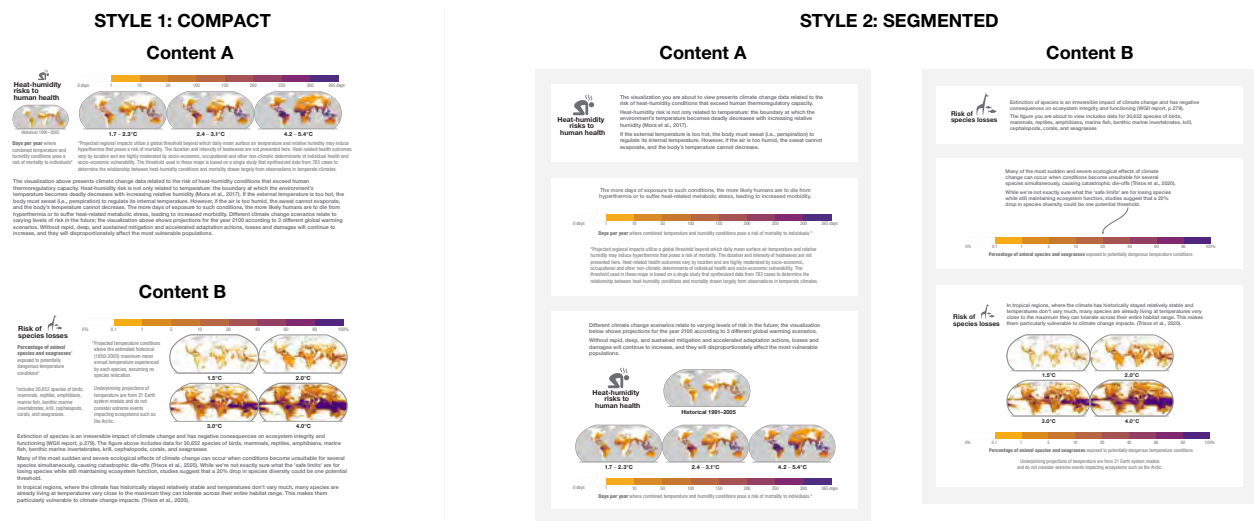


Figure 3. Each visualization content was adapted into two different presentation styles. Participants saw either A1 + B2, or A2 and B1. Source: Figure SPM3 in Calvin *et al.* (2023).

Presentation styles. I created two versions for each visualization: either *compact* (style 1) or *segmented* (style 2), as depicted in Figure 3. 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 I included the additional text below. In the segmented style, I divided the same content into three images: 1) the visualization’s title with the topic introduction part of the text, 2) the visualization’s color legend along with the portion of the text explaining the relevance of represented data w.r.t the topic, and 3) the visual representation, title, and legend, supplemented with the final part of the 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 Figure B1), or using arrow keys on their keyboard. The content did not differ between styles, except for the legend, which appeared twice in segmented versions. All images had a fixed width of 1100 pixels, ensuring

identical sizes of texts and maps across all stimuli.

Measures

Cognitive load. I measured cognitive load through subjective mental effort ratings, as it has shown to correlate with other cognitive load measures such as electroencephalogram (EEG) (Wang *et al.* 2020) while being “unintrusive” (Paas *et al.* 2003) and easy to implement. Paas (1992) developed a 9-point Likert scale, which allows 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 used different wording of that scale for the learners to disambiguate the different types of cognitive load in self-reports. Specifically, to assess **intrinsic load**, Ayres (2006) asked participants to rate how easy the learning content is from *very, very easy* to *very, very difficult*. Similarly, to measure **extraneous load**, Cierniak *et al.* (2009) asked participants to report the level of *easiness to learn with the material*, with responses ranging from *very, very easy* to *very, very difficult*. Wang *et al.* (2020) suggested **germane load** could be measured by asking participants to rate *how concentrated they are* during learning, from *very, very little* to *very, very much* (Salomon, 1984). In this study, I adapted from these works and I used items in Table C1. I fully labeled all Likert answer options (see Figure C1) to increase the reliability of collected data (DeCastellarnau, 2018).

Perceived readability. As I measured cognitive load from subjective ratings, I also collected subjective ratings of readability using PREVis (Cabouat *et al.* 2024b), a validated scale for evaluating perceived readability in data visualization. Respondents answer PREVis items on a 7-point bipolar agreement Likert scale, fully labeled as shown in Figure C2. The instrument includes 11 items across four scales (see Table C2) to measure how easily participants find they can *Understand* how to read a visualization (three items, $\alpha = 0.86$ and $\omega = 0.96$ in this survey), how clear they find the visualization’s *Layout* (three items, $\alpha = 0.80$ and $\omega = 0.93$ in this survey), how easily they feel they can *Read Data* values (three items, $\alpha = 0.84$ and $\omega = 0.96$ in this survey), and how easily they feel they can read *Data Features* such as extremums, patterns, or trends (two items, $\alpha = 0.67$ and $\omega = 0.86$ in this survey). Participants answered each scale on a separate screen in

random order, and items within each scale also appeared in a random position.

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: I collected all measures electronically using [Inria’s LimeSurvey platform](#).

Procedure

Each participant saw two visualizations: one was presented in three parts, according to the *segmenting* principle in Cognitive Theory of Multimedia Learning; and the other was presented as a single image. Both the content and order of appearance of these visualizations were randomized.

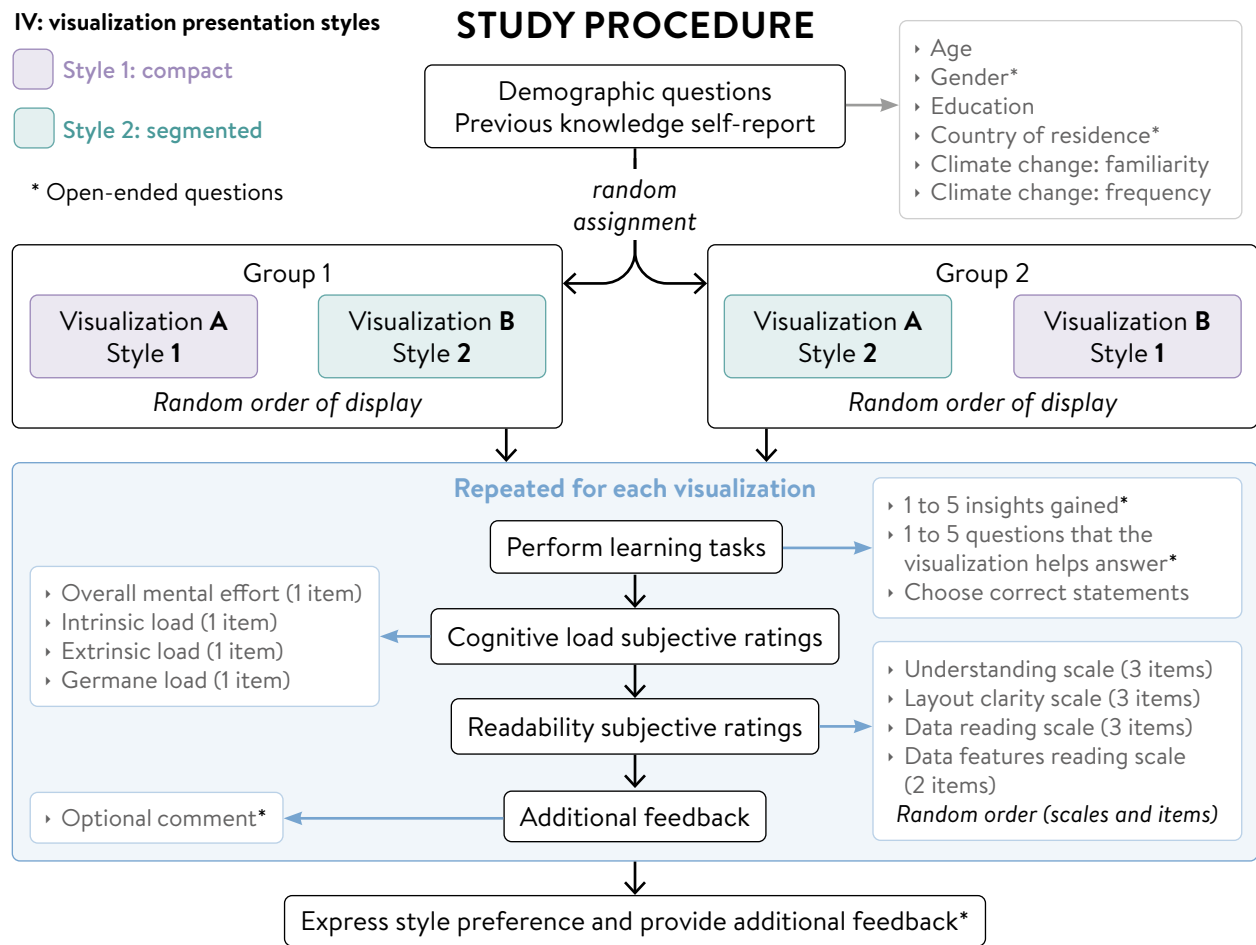


Figure 4. Overview of the main part of this study’s procedure. Participants repeated the blue segment for each of the two visualizations they saw.

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 I describe in [Figure 4](#). Just before seeing the first visualization, the survey displayed an information page providing some context in the form of the first sentence from Calvin *et al.* (2023), the IPCC report from which I extracted the visualizations. I also included a brief message acknowledging that emotions arising in response to the climate crisis can be uncomfortable. I 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 printout of the survey is available on OSF (osf.io/cj3pe).

Results

Collected data

Table 1. Collected ratings of cognitive load and readability for each Style \times Content condition.

Random group Stimulus	Group 1 ($N = 15$)				Group 2 ($N = 19$)			
	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
Intrinsic 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

[Table 1](#) above describes the collected cognitive load and perceived readability ratings across conditions. I describe collected style preferences in Appendix [Figure D2](#). 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.

Inferential Statistics

Readability and extraneous cognitive load correlation.

I created a repeated measures correlation matrix using the rmcrr 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. Repeated measures correlation matrix for subjective ratings of cognitive load (CL) and readability (PREVis). Significance labels: * $p < .05$, ** $p < .01$, *** $p < .001$

Figure 5 shows that readability ratings negatively correlated with extraneous load ratings. The correlation was moderate for the Data Features readability ratings ($r = -.50, p < 0.01$) and strong for the Understand ($r = -.77, p < 0.001$), Layout ($r = -.63, p < 0.001$), and Data Read ($r = -.64, p < 0.001$) ratings. I describe other correlations as exploratory analyses in Appendix F.

Effect of segmentation on cognitive load. A paired-samples t -test showed that learners did not experience a significantly higher overall cognitive load in the compact condition ($M = 5.59, SD = 1.5$) compared to the segmented condition ($M = 5.21, SD = 1.39$), $t(33) = 1.601, p = 0.060$. Due to this result and the small sample size for each randomized condition Group \times Order, I did not conduct further analyses planned in the pre-registration regarding this hypothesis.

I did, however, conduct a series of exploratory paired-samples t -tests to assess the effect of presentation style on other cognitive load measures (see Tables F1 and F2). Figure 6 shows the corresponding point estimates and confidence intervals. There was a significant difference for extraneous load between the compact and segmented styles, but not for intrinsic and germane load. As there were significant correlations between extraneous cognitive load and readability

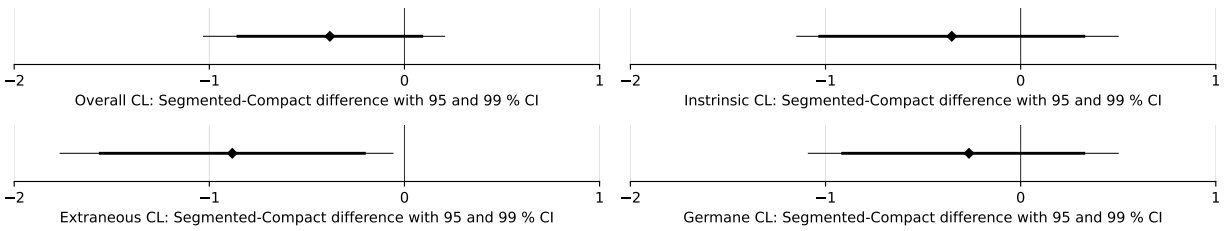


Figure 6. Within-participants differences in cognitive load ratings between *segmented–compact*. The point estimate is shown with with 95% (thick line) and 99% (thin line) confidence intervals (CI) Results are not statistically significant if the CI includes the null (0) value. In this study, only extraneous cognitive load measurements showed a significant difference, with a point estimate at -0.88.

measurements, I also conducted a series of exploratory paired-samples *t*-tests to assess the effect of style on perceived readability. There were significant differences for the Understand, Layout, and Data Read scales, but not for the Data Features scale (see details in [Tables F5 and F6](#)).

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. Three participants’ data were missing due to a technical issue, and two others did not explain their preferences in their answers. I conducted a thematic analysis on the remaining 29 comments. I provide the list of codes and themes in [Figure E1](#) and participants’ coded answers on the OSF repository (osf.io/pvrqjh).

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 takeaways up front and then expanded into details” (P7). Such comments resonate with Shneiderman’s (1996) 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: “I have to find where to start and do this work myself”

(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 [Figure D2](#)).

Discussion

My goal with this work was to assess the relevance of Cognitive Load Theory (CTL) and associated Cognitive Theory of Multimedia Learning (CTML) as a framework for designing and evaluating data visualizations. Findings from the correlation matrix in [Figure 5](#) support my 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).

Most correlations were strong except for the Data Features readability subscale. The reliability indices for this particular subscale were slightly weaker than those of the other subscales, which could have affected the strength of 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 readability subscales (see lists of PREVis items in [Table C2](#)). It is also possible that these items caused confusion among respondents as their wording is more complex than that of other PREVis items.

The Understand scale demonstrated a notably strong correlation with extraneous cognitive load ($r = 0.77$), suggesting that participants’ ability to comprehend *how to read* a data visualization might have greatly affected the amount of unnecessary cognitive load they experienced during learning tasks. This result supports the call from Stoiber *et al.* (2019) to develop visualization

onboarding practices, which the authors defined as “the process of supporting users on how to read, interpret, and extract information of visual representations of data.”

Overall, this study’s correlational findings add to existing empirical evidence supporting the relevance of the CLT framework for evaluating data visualizations (Anderson *et al.* 2011; Huang *et al.* 2009; Mautone & Mayer, 2007). My results specifically align with those of Huang *et al.* (2009), who suggested linking visualization task performance to cognitive load. While their study focused on overall cognitive load and performance metrics, I delved deeper into exploring links between visualization characteristics and various aspects of cognitive load, using subjective reports as a data collection method. My correlational analysis indicated that measures of extraneous cognitive load could be used as valid proxies for assessing the readability of data visualizations, for which we lack established measuring instruments (Cabouat *et al.* 2024a, 2024b). Further research should investigate the validity of other extraneous cognitive load measurements for evaluating readability—such as response time in a secondary visual task (DeLeeuw & Mayer, 2008), EEG measurements (Anderson *et al.* 2011), or eye activity (Padilla *et al.* 2020).

As a second hypothesis (H2), I expected learners’ overall mental effort to be lower in the segmented condition than in the compact condition. Although observations of overall cognitive load tended to align with this idea, the differences were not significant and H2 was not supported. In their meta-analysis, Rey *et al.* (2019) found that segmenting generally had a small effect on overall cognitive load, and no significant effect in the case of learner-paced segmentation. It is possible that the sample size in my current work was too small to detect a small effect; or that the segmenting effect does not reduce overall cognitive load when learning from data visualizations. Participants in my study mentioned the burden of manually navigating between parts of the visualization in the segmented condition. Researchers might prefer a system-paced segmentation in future work.

Another potential explanation for the lack of significant effect of segmentation on overall cognitive load is the *expertise reversal effect* in CTML (Kalyuga *et al.* 2003; Kalyuga, 2021). This principle explains how 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.* (2019)

identified prior domain knowledge as a moderating factor in their meta-analysis, although they found effects contradictory to what the theory predicted. The authors noted that the variance of learners' prior domain knowledge was very limited in their meta-analysis' pool of studies. Therefore, we do not have clear evidence of how learners' prior knowledge interacts with the segmenting effect. Unfortunately, my 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 aim to investigate the mediating effect of objectively assessed domain knowledge on cognitive load in visualization and expand from existing work from Richter *et al.* (2021) and Homer and Plass (2010) to examine how the reversal expertise effect applies in data visualization.

Although H2 was not supported, 20 participants out of 34 expressed a preference for the segmented version. These participants mentioned how segmentation helped them feel guided in processing the information piece by piece, while the compact version could be overwhelming at first. Exploratory analyses presented in [Appendix F](#) also showed a small but significant impact of style on extraneous cognitive load measures. These findings suggest that segmenting might be a suitable design approach for visualization onboarding (Stoiber *et al.* 2019, 2023).

A few methodological constraints in this study limit the conclusions that can be reached. One limitation was the inability to assess the reliability of subjective cognitive load measurements, as they all relied on single items. Unreliable measuring instruments can significantly impede our ability to observe a phenomenon (DeVellis & Thorpe, 2021). To address this issue, I recommend creating and validating cognitive load assessment scales for studying data visualization as well as other learning materials. The study's online setting also limited the reliability of cognitive load measurements. Some participants may not have completed the survey in one go, which could potentially impact the results, as subjective cognitive load ratings are highly sensitive to variations in mental workload (Paas *et al.* 2003). To further validate CLT in visualization contexts, future work should aim to ensure more controlled experimental conditions.

Conclusion

In this Master's Thesis, I propose an approach based on cognitive load to design and evaluate data visualizations. The results indicated that Cognitive Load Theory could serve as a valid framework for evaluating readability in data visualization contexts. However, when I applied the segmenting design principle derived from this theory to data visualizations, I did not find a significant impact on overall mental effort in learning tasks. While this manipulation did have a qualitative impact on participants' learning experiences, further work is required to address the limitations of my study. As is often the case both in data visualization and learning contexts, there is no one-size-fits-all design. Future research should aim to identify possible mediating factors for learning from data visualizations, such as the viewer's prior knowledge or visualization literacy. By integrating knowledge from human-computer interaction research, educational science, and cognitive psychology, we can begin to 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.

Acknowledgements

I would like to extend my heartfelt thanks to my supervisors for this work: Lorenzo Ciccione, Petra Isenberg, and Tobias Isenberg, for their invaluable contributions to my research. Their time, advice, and unwavering support have been essential to the success of this work. I am also deeply grateful to all the individuals in my professional and personal networks who helped disseminate information about this study, and to the participants who generously contributed their time and energy in responding to this call. I am grateful to Université Paris Cité, the EURIP graduate school, LPI, and our AIRE Master's team for providing us students with the opportunity to pursue this Master's program. I wish to thank all our teachers, and in particular Alexandra Laurent for her never-failing positive attitude and Macarena-Paz Celume for her two-year-long patient guidance in completing our thesis research projects. Finally, I am immensely thankful to all my friends and family for their attentiveness and reassuring words and to my classmates—especially the cooking sessions' crew—for showing incredible camaraderie. You made these two years of Master's studies a truly transformative experience.

Copyright and supplemental material

This document is shared under the CC BY 4.0 license. The visualizations visible in Figures 3, A1, A2, and B1 are the property of the IPCC and protected by intellectual property laws. The pseudonymized data collected from this study—stripped from demographic data and answers to learning tasks—is available on osf.io/pvrqh.

References

- Adar, E., & Lee, E. (2021). Communicative visualizations as a learning problem. *IEEE Transactions on Visualization and Computer Graphics*, 27(2), 946–956.
<https://doi.org/10/ghgt48>
- Anderson, E. W., Potter, K. C., Matzen, L. E., Shepherd, J. F., Preston, G. A., & Silva, C. T. (2011). A user study of visualization effectiveness using eeg and cognitive load. *Computer Graphics Forum*, 30(3), 791–800. <https://doi.org/10/excr9x>
- Ayres, P. (2006). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and Instruction*, 16(5), 389–400. <https://doi.org/10/b5xhfj>
- Bertin, J. (1967 (2011)). *Semiology of graphics: Diagrams, networks, maps*. ESRI Press.
- Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals*. Longmans, Green.
- Borgo, R., Abdul-Rahman, A., Mohamed, F., Grant, P. W., Reppa, I., Floridi, L., & Chen, M. (2012). An empirical study on using visual embellishments in visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12), 2759–2768. <https://doi.org/10/f4fqd3>
- Borkin, M. A., Bylinskii, Z., Kim, N. W., Bainbridge, C. M., Yeh, C. S., Borkin, D., Pfister, H., & Oliva, A. (2016). Beyond memorability: Visualization recognition and recall. *IEEE Transactions on Visualization and Computer Graphics*, 22(1), 519–528.
<https://doi.org/10/ggf5r3>
Dataset at <https://github.com/massvis/dataset>.
- Borkin, M. A., Vo, A. A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., & Pfister, H. (2013). What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics*, 19(12), 2306–2315. <https://doi.org/10/f5h3pd>
- Boy, J., Rensink, R. A., Bertini, E., & Fekete, J.-D. (2014). A principled way of assessing visualization literacy. *IEEE Transactions on Visualization and Computer Graphics*, 20(12), 1963–1972. <https://doi.org/10/f6qjv6>

- Brath, R. (1997). Metrics for effective information visualization. *Proceedings of VIZ '97: Visualization Conference, Information Visualization Symposium and Parallel Rendering Symposium*, 108–111. <https://doi.org/10/dm997r>
- Brehmer, M., & Munzner, T. (2013). A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12), 2376–2385. <https://doi.org/10/f5h3q4>
- Burns, A., Lee, C., On, T., Xiong, C., Peck, E., & Mahyar, N. (2022). From Invisible to Visible: Impacts of Metadata in Communicative Data Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 1–16. <https://doi.org/10/gtx89f>
- Burns, A., Xiong, C., Franconeri, S., Cairo, A., & Mahyar, N. (2020). How to evaluate data visualizations across different levels of understanding. *2020 IEEE Workshop on Evaluation and Beyond - Methodological Approaches to Visualization (BELIV)*, 19–28. <https://doi.org/10/kz34>
- Cabouat, A.-F., He, T., Cabric, F., Isenberg, T., & Isenberg, P. (2024a). Position paper: A case to study the relationship between data visualization readability and visualization literacy. *CHI 2024 - Workshop Toward a More Comprehensive Understanding of Visualization Literacy*. <https://hal.science/hal-04523790>
- Cabouat, A.-F., He, T., Isenberg, P., & Isenberg, T. (2024b). PREVis: Perceived readability evaluation for visualizations. *IEEE Transactions on Visualization and Computer Graphics* (conditionally accepted).
- Calvin, K., Dasgupta, D., Krinner, G., Mukherji, A., Thorne, P. W., Trisos, C., Romero, J., Aldunce, P., Barrett, K., Blanco, G., Cheung, W. W., Connors, S., Denton, F., Diongue-Niang, A., Dodman, D., Garschagen, M., Geden, O., Hayward, B., Jones, C., . . . Péan, C. (2023, July 25). *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 [Core Writing Team, H. Lee and J. Romero*

- (Eds.)). IPCC, Geneva, Switzerland. (First). Intergovernmental Panel on Climate Change (IPCC). <https://doi.org/10/k5jn>
- Camp, G., Surma, T., & Kirschner, P. A. (2021). Foundations of multimedia learning. In L. Fiorella & R. E. Mayer (Eds.), *The cambridge handbook of multimedia learning* (3rd ed., pp. 17–24). Cambridge University Press. <https://doi.org/10/gtx89p>
- Card, S. K., Mackinlay, J. D., & Shneiderman, B. (Eds.). (1999). *Readings in information visualization: Using vision to think* (1st edition). Morgan Kaufmann Publishers Inc.
- Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied*, 4(2), 75–100. <https://doi.org/10/ffbb2w>
- Ciccione, L. (2022, December). *The cognitive and neural bases of graph perception and comprehension* [Doctoral dissertation, PSL University].
- Cierniak, G., Gerjets, P., & Scheiter, K. (2009). Expertise reversal in multimedia learning: Subjective load ratings and viewing behavior as cognitive process indicators. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 31(31). Retrieved April 23, 2024, from <https://escholarship.org/uc/item/99r5z5fb>
- Correll, M., Bertini, E., & Franconeri, S. (2020). Truncating the y-axis: Threat or menace? *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–12. <https://doi.org/10/g9vj>
- Curcio, F. R. (1987). Comprehension of Mathematical Relationships Expressed in Graphs. *Journal for Research in Mathematics Education*, 18(5), 382–393. <https://doi.org/10/fjh8jc>
- DeCastellarnau, A. (2018). A classification of response scale characteristics that affect data quality: A literature review. *Quality & Quantity*, 52(4), 1523–1559. <https://doi.org/10/gdqv89>
- DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology*, 100(1), 223–234. <https://doi.org/10/cj4n4c>

- DeSanctis, G., & Jarvenpaa, S. (1985). An investigation of the "Tables versus Graphs" controversy in a learning environment. *ICIS 1985 Proceedings*.
- DeVellis, R. F., & Thorpe, C. T. (2021, September 16). *Scale development: Theory and applications*. SAGE Publications.
- Fansher, M., Adkins, T. J., Lalwani, P., Boduroglu, A., Carlson, M., Quirk, M., Lewis, R. L., Shah, P., Zhang, H., & Jonides, J. (2022). Icon arrays reduce concern over COVID-19 vaccine side effects: A randomized control study. *Cognitive Research: Principles and Implications*, 7(1), 38. <https://doi.org/10/grgcmc>
- Fox, A. R., & Hollan, J. D. (2023). Visualization psychology: Foundations for an interdisciplinary research program. In D. Albers Szafir, R. Borgo, M. Chen, D. J. Edwards, B. Fisher, & L. Padilla (Eds.), *Visualization psychology* (pp. 217–242). Springer International Publishing. <https://doi.org/10/gtgz76>
- Franconeri, S. L., Padilla, L. M., Shah, P., Zacks, J. M., & Hullman, J. (2021). The Science of Visual Data Communication: What Works. *Psychological Science in the Public Interest*, 22(3), 110–161. <https://doi.org/10/gnsdwm>
- Freedman, E. G., & Shah, P. (2002). Toward a model of knowledge-based graph comprehension. In M. Hegarty, B. Meyer, & N. H. Narayanan (Eds.), *Diagrammatic representation and inference* (pp. 18–30). Springer. <https://doi.org/10/fr56t9>
- Galesic, M., & Garcia-Retamero, R. (2011). Graph Literacy: A Cross-Cultural Comparison. *Medical Decision Making*, 31(3), 444–457. <https://doi.org/10/fqxhjn>
- Ghoniem, M., Fekete, J.-D., & Castagliola, P. (2005). On the readability of graphs using node-link and matrix-based representations: A controlled experiment and statistical analysis. *Information Visualization*, 4(2), 114–135. <https://doi.org/10/bncqv6>
- Harold, J., Lorenzoni, I., Shipley, T. F., & Coventry, K. R. (2020). Communication of ipcc visuals: Ipc authors' views and assessments of visual complexity. *Climatic Change*, 158(2), 255–270. <https://doi.org/10/gnffb6>

- Hegarty, M. (2011). The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science*, 3(3), 446–474. <https://doi.org/10/c3274w>
- Homer, B. D., & Plass, J. L. (2010). Expertise reversal for iconic representations in science visualizations. *Instructional Science*, 38(3), 259–276. <https://doi.org/10/dcfgcg>
- Huang, W., Eades, P., & Hong, S.-H. (2009). Measuring effectiveness of graph visualizations: A cognitive load perspective. *Information Visualization*, 8(3), 139–152. <https://doi.org/10/d3fs8b>
- Isenberg, P. (2021, November 30). *Micro visualizations* [thesis]. Université Paris-Saclay. Retrieved December 29, 2023, from <https://inria.hal.science/tel-03584024>
- Jarodzka, H. (2021). Research methods in multimedia learning. In L. Fiorella & R. E. Mayer (Eds.), *The cambridge handbook of multimedia learning* (3rd ed., pp. 41–54). Cambridge University Press. <https://doi.org/10/gtx89k>
- Kalyuga, S. (2021). The expertise reversal principle in multimedia learning. In L. Fiorella & R. E. Mayer (Eds.), *The cambridge handbook of multimedia learning* (3rd ed., pp. 171–182). Cambridge University Press. <https://doi.org/10/gtx893>
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38(1), 23–31. <https://doi.org/10/d7rktj>
- Kopp, T., Riekert, M., & Utz, S. (2018). When cognitive fit outweighs cognitive load: Redundant data labels in charts increase accuracy and speed of information extraction. *Computers in Human Behavior*, 86, 367–376. <https://doi.org/10/gg4zxw>
- Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology*, 3(3), 185–225. <https://doi.org/10/cvdzr5>
- Krosnick, J. (2017, October 23). Questionnaire design. <https://doi.org/10/gqcw3j>
- Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11(1), 65–100. <https://doi.org/10/d9pkk4>
- Lee, S., Kim, S.-H., Hung, Y.-H., Lam, H., Kang, Y.-A., & Yi, J. S. (2015). How do people make sense of unfamiliar visualizations?: A grounded model of novice’s information visualization

- sensemaking. *IEEE Transactions on Visualization and Computer Graphics*, 22(1), 499–508. <https://doi.org/10/gfw4vs>
- Lee, S., Kim, S.-H., & Kwon, B. C. (2017). Vlat: Development of a visualization literacy assessment test. *IEEE Transactions on Visualization and Computer Graphics*, 23(1), 551–560. <https://doi.org/10/f92d38>
- Lee-Robbins, E., He, S., & Adar, E. (2022). Learning objectives, insights, and assessments: How specification formats impact design. *IEEE Transactions on Visualization and Computer Graphics*, 28(1), 676–685. <https://doi.org/10/gtx89j>
- Mautone, P. D., & Mayer, R. E. (2007). Cognitive aids for guiding graph comprehension. *Journal of Educational Psychology*, 99(3), 640–652. <https://doi.org/10/dc3dcr>
- Mayer, R. E. (Ed.). (2001a). *Multimedia learning*. Cambridge University Press. <https://doi.org/10/gtx89q>
- Mayer, R. E. (2001b). Multimedia principle. In *Multimedia learning* (pp. 63–80). Cambridge University Press. <https://doi.org/10/gtx89r>
- Mayer, R. E. (2021). Cognitive theory of multimedia learning. In L. Fiorella & R. E. Mayer (Eds.), *The cambridge handbook of multimedia learning* (3rd ed., pp. 57–72). Cambridge University Press. <https://doi.org/10/gtx89t>
- Mayer, R. E., & Fiorella, L. (Eds.). (2021a). *The cambridge handbook of multimedia learning* (3rd ed.). Cambridge University Press. <https://doi.org/10/gtx89s>
- Mayer, R. E., & Fiorella, L. (2021b). Principles for managing essential processing in multimedia learning: Segmenting, pre-training, and modality principles*. In L. Fiorella & R. E. Mayer (Eds.), *The cambridge handbook of multimedia learning* (3rd ed., pp. 243–260). Cambridge University Press. <https://doi.org/10/gtx89z>
- McCrudden, M. T., & Van Meter, P. N. (2021). Multimedia learning with visual displays. In L. Fiorella & R. E. Mayer (Eds.), *The cambridge handbook of multimedia learning* (3rd ed., pp. 510–520). Cambridge University Press. <https://doi.org/10/gtx89w>

- McMahon, R., Stauffacher, M., & Knutti, R. (2015). The unseen uncertainties in climate change: Reviewing comprehension of an ipcc scenario graph. *Climatic Change*, *133*(2), 141–154. <https://doi.org/10/f7v2bq>
- Munzner, T. (2014, December 1). *Visualization analysis and design* (1st edition). CRC Press.
- Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*(4), 429–434. <https://doi.org/10/c74hkc>
- Paas, F., & Sweller, J. (2021). Implications of cognitive load theory for multimedia learning. In L. Fiorella & R. E. Mayer (Eds.), *The cambridge handbook of multimedia learning* (3rd ed., pp. 73–81). Cambridge University Press. <https://doi.org/10/gtx89n>
- Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*(1), 63–71. <https://doi.org/10/bznn3t>
- Padilla, L. M., Castro, S. C., Quinan, P. S., Ruginski, I. T., & Creem-Regehr, S. H. (2020). Toward objective evaluation of working memory in visualizations: A case study using pupillometry and a dual-task paradigm. *IEEE Transactions on Visualization and Computer Graphics*, *26*(1), 332–342. <https://doi.org/10/ghs95w>
- Pandey, A. V., Rall, K., Satterthwaite, M. L., Nov, O., & Bertini, E. (2015). How deceptive are deceptive visualizations? an empirical analysis of common distortion techniques. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 1469–1478. <https://doi.org/10/gfz5js>
- Pinker, S. (1990). A theory of graph comprehension. In *Artificial intelligence and the future of testing* (pp. 73–126). Lawrence Erlbaum Associates, Inc.
- Quadri, G. J., & Rosen, P. (2022). A survey of perception-based visualization studies by task. *IEEE Transactions on Visualization and Computer Graphics*, *28*(12), 5026–5048. <https://doi.org/10/gr6323>

- Quadri, G. J., Wang, A. Z., Wang, Z., Adorno, J., Rosen, P., & Szafir, D. A. (2024, February 23). *Do you see what i see? a qualitative study eliciting high-level visualization comprehension*. arXiv: 2402.15605 [cs]. <https://doi.org/10/gtx892>
- Rey, G. D., Beege, M., Nebel, S., Wirzberger, M., Schmitt, T. H., & Schneider, S. (2019). A meta-analysis of the segmenting effect. *Educational Psychology Review*, 31(2), 389–419. <https://doi.org/10/gtx89x>
- Richter, J., Wehrle, A., & Scheiter, K. (2021). How the poor get richer: Signaling guides attention and fosters learning from text-graph combinations for students with low, but not high prior knowledge. *Applied Cognitive Psychology*, 35(3), 632–645. <https://doi.org/10/gndvdd>
- Salomon, G. (1984). Television is "easy" and print is "tough": The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of Educational Psychology*, 76(4), 647–658. <https://doi.org/10/bq9z6d>
- Sánchez-Holgado, P., Arcila-Calderón, C., & Frías-Vázquez, M. (2023). The effect of interest and attitude on public comprehension of news with data visualization. *Frontiers in Communication*, 8. <https://doi.org/10/gtx89g>
- Scheiter, K., & Gerjets, P. (2007). Learner control in hypermedia environments. *Educational Psychology Review*, 19(3), 285–307. <https://doi.org/10/c9kw9g>
- Schnotz, W., & Kürschner, C. (2008). External and internal representations in the acquisition and use of knowledge: Visualization effects on mental model construction. *Instructional Science*, 36(3), 175–190. <https://doi.org/10/c6gxbw>
- Shah, P., & Freedman, E. G. (2011). Bar and line graph comprehension: An interaction of top-down and bottom-up processes. *Topics in Cognitive Science*, 3(3), 560–578. <https://doi.org/10/frczb8>
- Shah, P., & Hoeffner, J. (2002). Review of Graph Comprehension Research: Implications for Instruction. *Educational Psychology Review*, 14(1), 47–69. <https://doi.org/10/cq7tq5>

- Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. *Proceedings 1996 IEEE Symposium on Visual Languages*, 336–343. <https://doi.org/10/fwdq26>
- Stoiber, C., Grassinger, F., Pohl, M., Stitz, H., Streit, M., & Aigner, W. (2019). *Visualization onboarding: Learning how to read and use visualizations* (OSF Preprints). <https://doi.org/10/gh38zd>
- Stoiber, C., Wagner, M., Grassinger, F., Pohl, M., Stitz, H., Streit, M., Potzmann, B., & Aigner, W. (2023). Visualization onboarding grounded in educational theories. In D. Albers Szafir, R. Borgo, M. Chen, D. J. Edwards, B. Fisher, & L. Padilla (Eds.), *Visualization psychology* (pp. 139–164). Springer. <https://doi.org/10/gtjwqn>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. <https://doi.org/10/cmgsv8>
- Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development*, 68(1), 1–16. <https://doi.org/10/ggjjgpr>
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer. <https://doi.org/10/ddz49m>
- Thudt, A., Walny, J., Perin, C., Rajabiyazdi, F., MacDonald, L., Vardeleon, R., Greenberg, S., & Carpendale, S. (2016). Assessing the readability of stacked graphs. *Proceedings of the 42nd Graphics Interface Conference*, 167–174. <https://doi.org/10/gtgz8d>
- Wang, J., Antonenko, P., Keil, A., & Dawson, K. (2020). Converging subjective and psychophysiological measures of cognitive load to study the effects of instructor-present video. *Mind, Brain, and Education*, 14(3), 279–291. <https://doi.org/10/ggwqgp>
- Wong, Y. L., Madhavan, K., & Elmqvist, N. (2018). Towards characterizing domain experts as a user group. *2018 IEEE Evaluation and Beyond - Methodological Approaches for Visualization (BELIV)*, 1–10. <https://doi.org/10/gtx89h>

- Yoghourdjian, V., Yang, Y., Dwyer, T., Lawrence, L., Wybrow, M., & Marriott, K. (2021). Scalability of network visualisation from a cognitive load perspective. *IEEE Transactions on Visualization and Computer Graphics*, 27(2), 1677–1687. <https://doi.org/10/ghs96j>
- Zeng, Z., & Battle, L. (2023). A review and collation of graphical perception knowledge for visualization recommendation. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–16. <https://doi.org/10/kz36>
- Zhu, Y. (2007). Measuring effective data visualization. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, N. Paragios, S.-M. Tanveer, T. Ju, Z. Liu, S. Coquillart, C. Cruz-Neira, T. Müller, & T. Malzbender (Eds.), *Advances in visual computing* (pp. 652–661). Springer. <https://doi.org/10/cxjb29>

Appendix

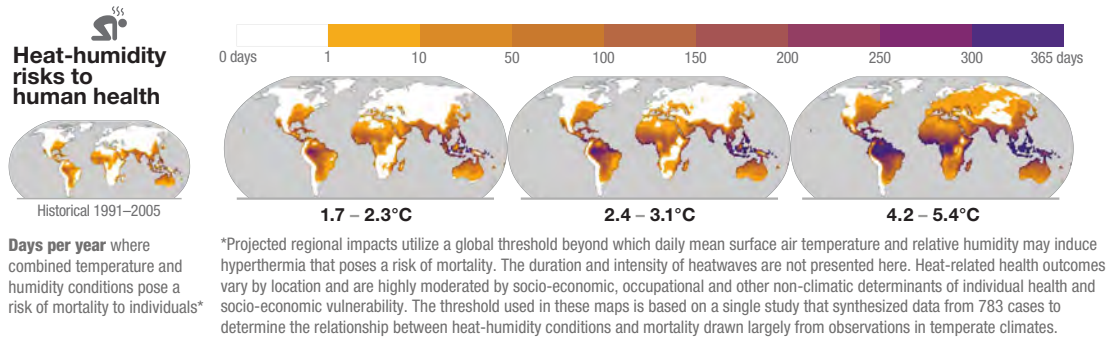
Table of Contents

Stimuli	ii
Survey screenshot	iii
Measuring instruments	iv
Cognitive load	iv
Perceived readability (PREVis)	v
Collected data	vi
Demographics table	vi
Previous exposure to climate change	vii
Data exclusion	viii
Style preferences	x
Qualitative analysis of preference motivations	xi
Inferential data analysis	xii
Repeated measures correlations	xii
Paired-samples ttests	xiii

Appendix A

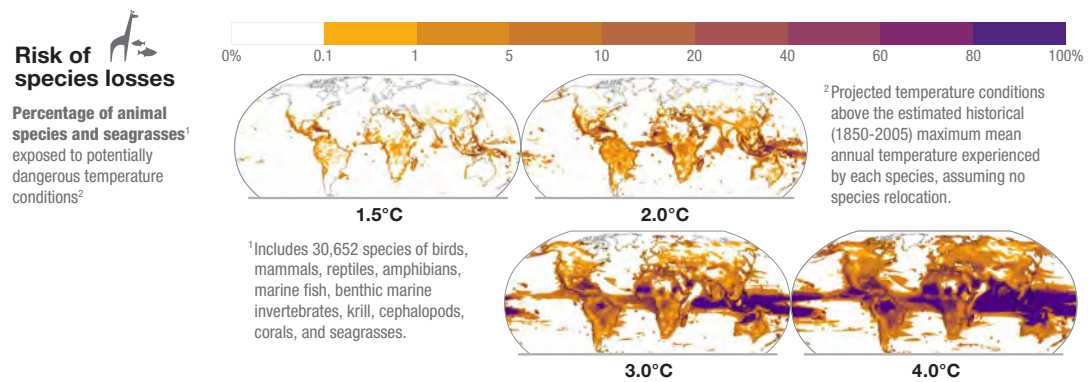
Stimuli

Figure A1 and Figure A2 below present the study’s stimuli contents. Upper parts correspond to the original representation from Figure SPM3 in Calvin *et al.* (2023). The text block below corresponds to the additional explanations I created based on the report (Calvin *et al.* 2023) and underlying studies.



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 A1. Content A presented in Style 1. Content remained the same in Style 2 but was segmented as described in the Method section . Source: Figure SPM3 in Calvin *et al.* (2023).



Underpinning projections of temperature are from 21 Earth system models and do not consider extreme events impacting ecosystems such as the Arctic.

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

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.

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).

Figure A2. Content B presented in Style 1. Content remained the same in Style 2 but was segmented as described in the Method section . Source: Figure SPM3 in Calvin *et al.* (2023).

Appendix B

Survey screenshot

Please read the following text and data visualization.

Imagine that you find the visualization below in an online news article.
You are interested in learning about the topic, and decide to study it with attention, along with the accompanying caption and labels.

Scrolling down, you will find a visualization comprising three (3) successive views. Only one view is visible at a time.
You can use the round buttons or the arrows on your keyboard to navigate between the views.

Take as long as you want to read and learn information from the visualization and accompanying texts.

You may then scroll further down to answer the first question. Please note that you can only continue to the next screen once you have seen all three (3) parts of the visualization.

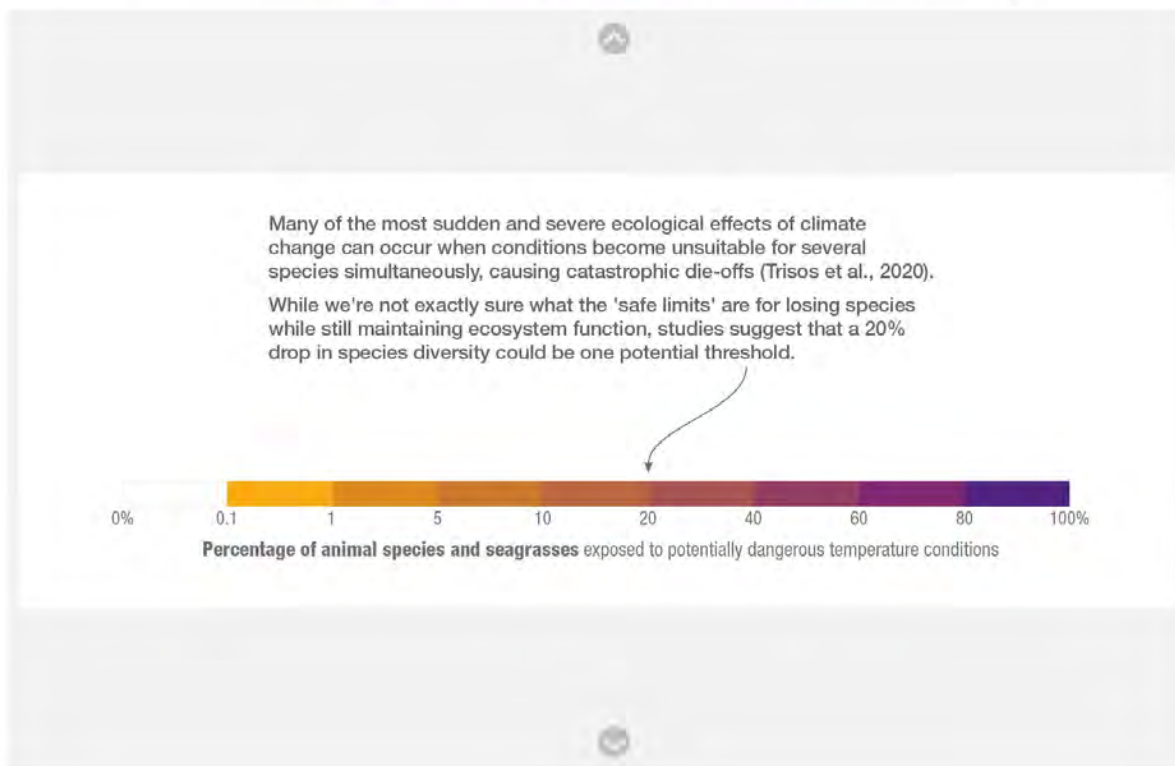


Figure B1. Example screenshot for the presentation of visualization B2, i. e., content B (risk of species loss) in style 2 (segmented in three parts). The image in the figure was the second segment out of three. The screenshot also shows the survey's accompanying text above the visualization and up/down buttons available to navigate between visualization segments.

Appendix C

Measuring instruments

Cognitive load

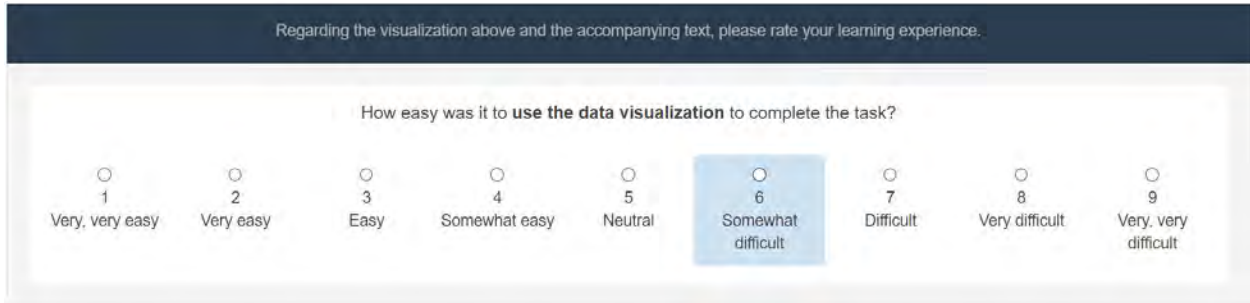


Figure C1. Screenshot of the “Extraneous load” rating item and answer options.

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

Table C1. Questions and Likert extremum labels used to collect subjective cognitive load ratings. All points were numbered and labeled, as shown in [Figure C1](#) above.

Perceived readability (PREVis)

Regarding the visualization above, to what degree do you agree or disagree with the following statement:

I don't find this visualization messy

Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree Strongly agree

I don't find this visualization crowded

Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree Strongly agree

I don't find parts of the visualization distracting

Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree Strongly agree

Previous Next

Figure C2. Example screenshot of the “Layout” PREVis scale.

Table C2. 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 2 - Disagree 3 - Slightly disagree 4 - Neutral 5 - Slightly agree 6 - Agree 7 - Strongly agree
	I can easily understand how the data is represented in this visualization	
	I can easily understand this visualization	
Layout	I don't find this visualization messy	
	I don't find this visualization crowded	
	I don't find parts of the visualization distracting	
Data features	I find data features (for example, a minimum, or an outlier, or a trend) visible in this visualization	
	I can clearly see data features (for example, a minimum, or an outlier, or a trend) in this visualization	
Data read	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	

Appendix D
Collected data

Demographics table

Table D1. *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

Previous exposure to climate change

Table D2. *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

Table D3. *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

Data exclusion

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, I had included the image from [Figure D1](#) 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 D1. I 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. I conducted data analyses using only information from these 34 participants and I deleted the other entries. A few hours after starting the data collection, I checked the first answers using Python. I 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). I 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, I retained their data for the main inferential analyses and considered them as missing data for preference analysis (see [Figure D2](#)).

Attention checks

As pre-registered (osf.io/ptsne), I 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 sensible—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*. I 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. I am 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 my exclusion criteria. Therefore, I 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, I did not exclude a single participant from the data analysis based on answer quality checks.

Style preferences

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

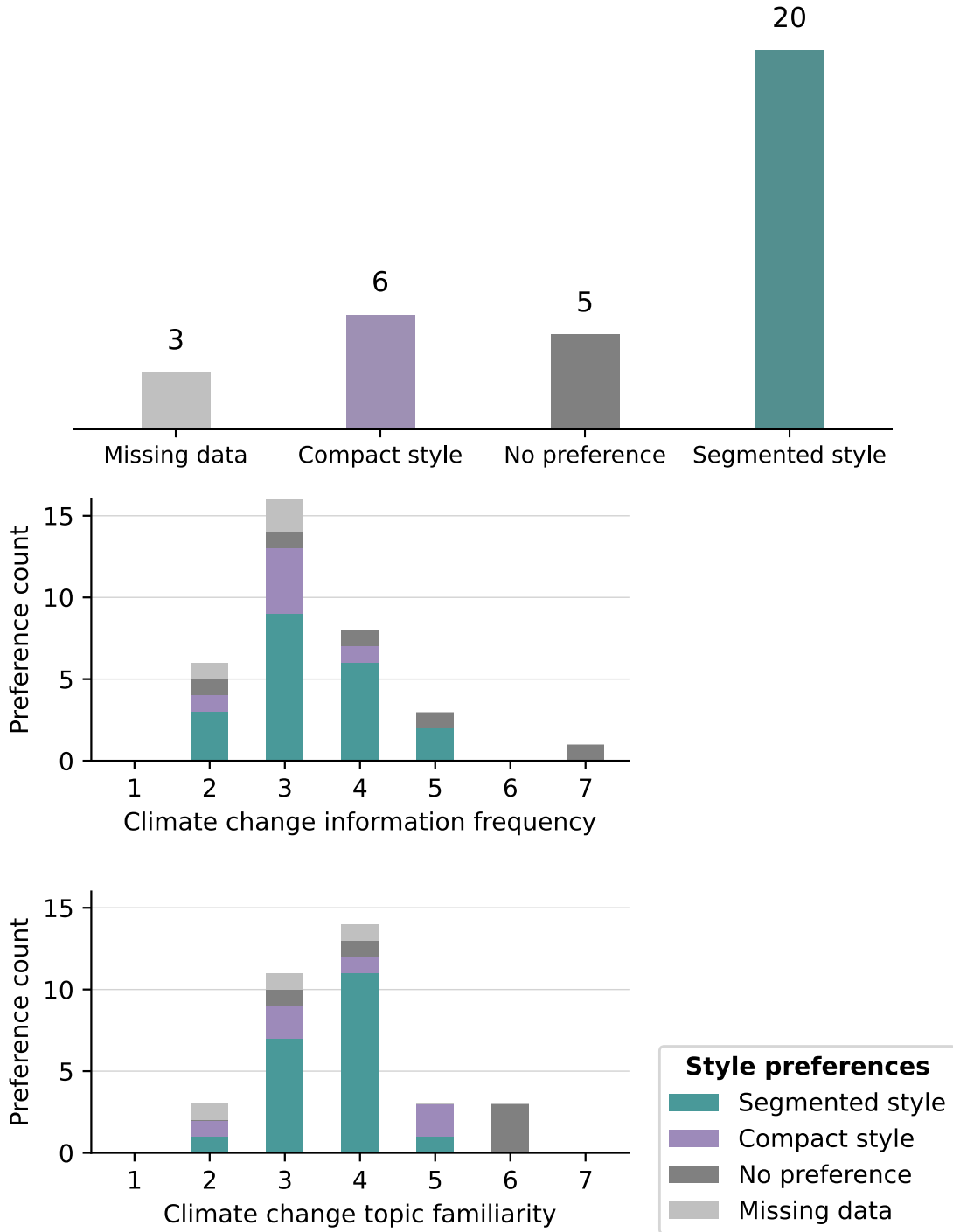


Figure D2. Top: style preferences distribution among all participants. Middle: style preferences by topic familiarity. Bottom: style preferences by exposure frequency.

Appendix E

Qualitative analysis of preference motivations

I conducted a thematic analysis of participants' comments regarding their style preference to identify what motivated their choice. Collected comments and their corresponding codes are available on osf.io/c73ys. Below is the summary of the codes and themes I identified.

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 takeaways 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 E1. Codes and themes from the participants' comments on their choice of preferred style.

Appendix F
Inferential data analysis

Repeated measures correlations

Correlation matrix

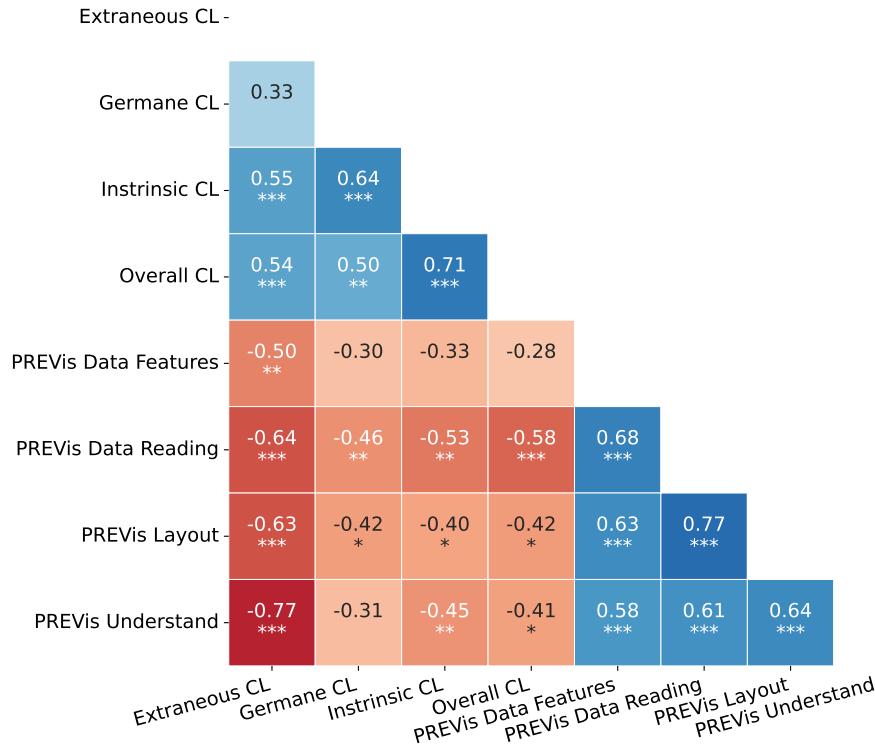


Figure F1. Repeated measures correlation matrix for subjective ratings of cognitive load (CL) and readability (PREVis) generated using the *rmcorr* package in R. Significance labels: * $p < .05$, ** $p < .01$, *** $p < .001$

Figure F1 above is the reproduction of Figure 5. As presented in the Results Table , it reveals that perceived readability ratings negatively correlated with subjective ratings of extraneous cognitive load. Further exploration of this matrix shows that the Data Read 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 scale 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 ($r = .33, p > 0.05$).

Paired-samples ttests

Effect of presentation style on cognitive load ratings

Table F1 provides descriptive values for cognitive load measurements across styles. As reported in the Results section, I 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 F2 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.

	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 F1. Descriptive statistics for cognitive load ratings between compact presentation style (style 1) and segmented presentation style (style 2).

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

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

Table F2. Paired sample *t*-tests for cognitive load ratings between compact presentation style (style 1) and segmented presentation style (style 2).

Having found a significant difference in extraneous cognitive load between compact and segmented styles, I 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. I 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 Tables F3 and F4 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.

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η_p^2
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 F3. Repeated Measures ANOVA: Within-subjects effects of style, 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_p^2
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

Table F4. Repeated Measures ANOVA: Between subjects effects of order of display (*vis_order*) and visualization content (*rand_group*) on extraneous cognitive load.

Effect of presentation style on readability ratings

Table F5 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, I conducted a series of exploratory paired-samples *t*-tests to assess the effect of presentation style on readability, as shown in Table F6. There were significant differences between compact and segmented styles for the Understand, Layout, and Data Read readability scales but not for the Data Features scale.

	N	Mean	Median	SD	SE
Style1 Understand	34	4.86	5.00	1.51	0.26
Style2 Understand	34	5.56	5.67	1.20	0.21
Style1 Layout	34	4.04	3.83	1.68	0.29
Style2 Layout	34	5.21	6.00	1.56	0.27
Style1 DataRead	34	4.57	5.00	1.62	0.28
Style2 DataRead	34	5.20	5.50	1.31	0.22
Style1 DataFeat	34	4.81	5.00	1.66	0.28
Style2 DataFeat	34	5.25	5.50	1.49	0.26

Table F5. Descriptive statistics for perceived readability ratings between compact presentation style (style 1) and segmented presentation style (style 2).

			statistic	df	p	Effect Size	95% Confidence Interval	
							Lower	Upper
Style1 Understand	Style2 Understand	Student's <i>t</i>	-2.69	33.0	0.006	Cohen's <i>d</i>	-0.461	-0.812 -0.1041
Style 1 Layout	Style2 Layout	Student's <i>t</i>	-3.64	33.0	< .001	Cohen's <i>d</i>	-0.625	-0.989 -0.2524
Style 1 DataRead	Style2 DataRead	Student's <i>t</i>	-2.19	33.0	0.018	Cohen's <i>d</i>	-0.376	-0.721 -0.0252
Style1 DataFeat	Style2 DataFeat	Student's <i>t</i>	-1.58	33.0	0.062	Cohen's <i>d</i>	-0.270	-0.611 0.0740

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

Table F6. Paired sample *t*-tests for perceived readability ratings between compact presentation style (style 1) and segmented presentation style (style 2).

