House Advantage or House of Cards? Stacking the Deck for Data Videos Leads to Null Results

Jen Rogers
Department of Computer Science
Tufts University
Boston, Massachusetts, USA
jen@cs.tufts.edu

Leixian Shen
Department of Computer Science and Engineering
The Hong Kong University of Science and Technology
Hong Kong, China
lshenaj@connect.ust.hk

Ab Mosca
Department of Computer Science
Smith College
Northampton, Massachusetts, USA
amosca@smith.edu

Evan Peck
Department of Information Science
University of Colorado Boulder
Boulder, Colorado, USA
evan.peck@colorado.edu

Mingwei Li
Department of Computer Science
Tufts University
Boston, Massachusetts, USA
mingwei.li@tufts.edu

Anzu Hakone
Department of Computer Science
Tufts University
Medford, Massachusetts, USA
anzu.hakone@gmail.com

Kristi Potter
Computational Science
National Renewable Energy Laboratory
Golden, Colorado, USA
kristi.potter@nrel.gov

Remco Chang
Department of Computer Science
Tufts University
Medford, Massachusetts, USA
remco@cs.tufts.edu

Abstract

Videos are becoming a ubiquitous means of sharing information on social media platforms. In response, data videos—short clips combining visualization with dynamic storytelling, audio descriptions, and spatial referencing—have gained popularity for communicating data. These affordances suggest that data videos might communicate data patterns, trends, and concepts more effectively than static visualizations, enhancing comprehension. However, existing research has not systematically tested this claim. To address this gap, we conducted two controlled studies to measure comprehension differences between data videos and static visualizations. Despite leveraging visual cues and audio explanations, no data video led to significantly better comprehension than an analogous static visualization. Our results suggest data videos are not categorically better and that future research should examine the tradeoffs between their engagement benefits and costs.

CCS Concepts

• **Human-centered computing** → **Information visualization**.

Keywords

Data Video, Static Visualization, Comprehension

ACM ISBN 979-8-4007-1395-8/25/04
https://doi.org/10.1145/3706599.3720256

ACM Reference Format:
Jen Rogers, Leixian Shen, Ab Mosca, Evan Peck, Mingwei Li, Anzu Hakone, Kristi Potter, and Remco Chang. 2025. House Advantage or House of Cards? Stacking the Deck for Data Videos Leads to Null Results. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3706599.3720256

1 Introduction

Just shy of 20 million views, the late Hans Rosling's talk on global health; "The Best Stats You've Ever Seen," is widely cited in the canon of data visualization. The animated bubbles, accompanied by Rosling's charismatic play-by-play commentary, achieve an energy inconceivable from a static chart. This example [32] and others like it [43], stand out for their combination of engaging narrative with dynamic visuals, making data accessible and engaging to a wider audience - important goals in data video research [4] and narrative visualization more broadly.

Data videos, a subset of narrative data visualization [10], combine animated visualizations with a narrative to communicate data-driven insights [10, 12, 38, 39]. As in Rosling's example, data racing across the screen can capture attention and related work has indicated user preference toward data videos over static charts [4, 20, 46]. But are data videos more effective than static visualizations? Despite their growing popularity, existing research has not conclusively articulated the benefits of data videos over static visualization in enhancing data comprehension.

The lack of consensus on the effectiveness of data videos stems from various factors. First, there is no consensus on the definition and use of the term "data video" in academic studies. For example, both Hans Rosling's data visualization talks and animated D3 visualizations have been labeled as data videos [8, 32]. This ambiguity makes it difficult to assess which elements—such as actors, narrative structure, narration, animation, or visual design—impact their effectiveness.

Second, there is a lack of standardized metrics for evaluating the benefits of data videos. As a relatively new medium, the research community has not yet established a common framework for assessing their relevance and effectiveness. Although data videos are perceived as successful in facilitating data communication and increasing viewer engagement [3–5, 12, 34], it is unclear whether this success is due to the growing familiarity with video as a format [25] or inherent advantages of data videos over traditional static visualizations.

Motivated by this gap, we conducted two studies investigating whether data videos are more effective than static visualizations for comprehension. Although we designed our studies to "stack the deck," highlighting the potential advantages of data videos, our findings suggest that designers should not expect clear benefits from converting static visualizations into data videos. Instead, they must consider the medium's affordances, audience, and the visualization's characteristics. Our findings also suggest that the research community should evaluate data videos beyond user preference, including their long-term effects on social engagement and learning outcomes.

The primary contribution of this work is an initial empirical examination of the effectiveness of data videos, which results in defined limitations of the data video medium and areas for future work. Our secondary contribution is a categorization of data videos with which we situated our experiments and related studies. We suggest that designers consider the context of data video use and call for future research on their long-term effects on learning and engagement beyond immediate comprehension.

2 Related Work

Data Video Definitions: *Data video* is a narrative format presenting a key data-driven message through a sequence of events, featuring at least one visualization [3, 38]. This broad definition covers various media types that use data video elements to tell stories—such as visualizations, real-world footage, animations, and voice-overs. Examples range from Hans Rosling's dynamic presentations to simple animations. For this study, we organized these examples into four categories (C1-C4): ***C1: Animated Unit*** refers to a single animation effect applied to specific visual elements, akin to a standalone animation. These units primarily capture attention by adding motion to static elements, but they do not convey a complete narrative [8, 15, 18, 19, 23, 47]. ***C2: Animated Narrative Visualization without Audio Narration*** consists of a sequence of animation units that together form a complete story, similar to a single or multi-slide presentation without voiceover. ***C3: Animated Narrative Visualization with Audio Narration*** builds on the previous category by incorporating audio narration, adding an auditory channel alongside the visual stimuli. This format is similar to a recorded slide presentation with voiceover [14, 20, 33]. The audio can be human-recorded [16], or generated from text [28, 35–37, 39, 44, 46]. ***C4: Context-Enhanced Animated Visualization with Audio Narration*** expands on C3 by incorporating contextual embellishments such as real-world scenes, graphic illustrations, or the presence of a presenter. Current research on this category mainly focuses on enriching sports videos with visualizations[11, 21, 40, 45], and recording presenter's gestures overlaid on charts [16].

Our study focuses on **C3**, which integrates both visual and auditory components, allowing a more precise evaluation of data video effectiveness compared to static visualizations. This category provides a balanced, controlled setting to explore how combining audio narration with animation improves user comprehension.

Conflicts in Data Video Evaluation: A common measure to evaluate visualizations is the accuracy of tasks or questions related to the visualization Generally, these studies find null ([4], [20], [9], [27], [6]) or negative ([31]) results for fact-based comprehension when comparing data videos to static or interactive visualizations. Amini et al. [4] examined the impact of pictographs and animation on user task performance (C2). Kong et al. [20] studied the effect of 8 different visual cues accompanied by audio narrations on comprehension(C3). Both found no statistically significant differences in comprehension across the different conditions. While Robertson et al. [31] found animation trend visualizations (C2) to lead to worse task accuracy in than small multiples [31], Brehmer et al.'s [9] focus on mobile devices and Archembault et al.'s analsys of mental map preservation [6] found no statistically significant differences.

However, another set of studies found the benefits of animation on learning[7, 13, 22, 26, 41]. For example, Obie et al.'s [26] study found a type of comprehension in which data videos with audio narration (C3) outperformed static visualization. They tested comprehension of "facts", i.e. participants' ability to read facts from the static or animated visualization, and comprehension of "value", i.e. participants' ability to understand the take-away message of the visualization. While they found no significant difference in "fact" comprehension, they found animated visualizations were significantly better than static ones for "value" comprehension. A meta-analysis of 61 between-group experiments found a small but overall positive effect ($g = 0.226$) of animation over static visualizations [7]. Studies also show that system-paced animated materials and more realistic or schematic forms of animation can yield stronger benefits than purely abstract visuals [7]. These nuances suggest that while animation alone does not guarantee improved fact-based comprehension, carefully chosen design elements and contexts can enhance its effectiveness.

Work in this area reveals a mismatch between subjective and objective measures of comprehension. Two studies [27] [31] found null or negative results on task accuracy, but positive results from subjective ratings of understandability and helpfulness. For example, Ying et al. [46] found animated charts (C2) and live videos (C3) were scored higher for understandability than static visualizations by study participants. Other studies discovered significantly higher user engagement [4] and higher self-reported measures of confidence [9] .

However, even subjective findings are not consistent. Robertson et al. [31] compared static, animated, and trace videos (animated videos with narration) and found participants rated animation as more helpful than traces for understanding. In a similar vein, Kong et al. [20] found no significant differences in recall across cued and static visualizations.

Figure 1: High-level structure of experiments. Each pre and post-study questionnaire was identical for both experiments. The main task for each experiment that differed from one another is highlighted in blue.

3 Study Motivation and Hypotheses

Our research investigates when data videos enhance comprehension and which factors contribute to their effectiveness. While literature shows a preference for data videos [4], subjective ratings of preference and engagement do not always match objective comprehension and accuracy measures. We designed our experiments to isolate factors unique to data videos as defined for C3, such as narration, animation, and visual cues, while excluding factors like actors and narrative structures. Our study examines the trade-offs between data videos and static visualizations, focusing on how animation and narration in data videos affect comprehension, cognitive load, and visual literacy compared to static visualizations with identical informational content. **We propose the following hypotheses:**

H(A) – Accuracy: We hypothesize that data videos improve accuracy in interpreting unfamiliar visualizations. This is due to spatial referencing, which links narration to visual elements, helping viewers process information efficiently and reducing errors. In Experiment 1 (Sec. 4), we tested this using a custom, unfamiliar visualization to assess participants' comprehension accuracy across two visualization types.

H(B) – Subjective Rating: We hypothesize that data videos reduce cognitive effort due to dual-modality (visual and auditory) presentation and spatial referencing, which aid comprehension. Previous research supports multi-modal information presentation (e.g., [14, 20, 31, 33]). To test this, we surveyed participants' engagement and task load using the engagement questionnaire from Amini et al. [4] and the NASA TLX questionnaire [17].

H(C) – Visual Literacy: We hypothesize that data videos require lower visual literacy than static visualizations. The use of multiple modalities and spatial referencing helps viewers with lower visual literacy learn more effectively, making data videos more accessible. We tested this using Mini-VLAT questionnaires by Pandey et al. [29] in both experiments.

H(D) – Speed: We hypothesize that static visualizations will be consumed more quickly than data videos because they allow readers to control their time and focus. Data videos follow a fixed linear progression, which may take longer. We recorded completion times in both experiments to test this hypothesis.

4 Experiment 1: Bespoke Visualization

Experiment 1 was motivated by the postulation that data videos are better at explaining bespoke visualizations than static visualizations, leading to a higher comprehension of visual encoding. **We hypothesized that data videos would improve understanding of unfamiliar visual representations and data misaligned with viewers' mental models.** Specifically, we test the following null hypotheses: $\mathbf{H_0}$**(1A)**: There is no difference in visual encoding comprehension between participants in the two conditions. $\mathbf{H_0}$**(1B)**: There is no difference in participants' self-reported cognitive efforts between conditions. $\mathbf{H_0}$**(1C)**: There is no difference in the visualization literacy required to comprehend data video and static visualizations. $\mathbf{H_0}$**(1D)**: There is no difference in visualization viewing time between the data video and static visualization conditions.

4.0.1 *Set-Up.*

We ran a between-subjects A-B (data video vs. static visualization) study on Prolific [30]. Participants were paid \$8 per hour for their time. We required them to be fluent in English and have working audio. Figure 1 shows the flow of the experiments, with the main task highlighted in blue. We leveraged related work, using an engagement questionnaire [4], the NASA TLX [17], and Mini-VLAT [29]. We recorded the time taken to complete the entire experiment, their time with the stimulus, and their time to complete the task questions.

4.0.2 *Stimuli.*

Experiment 1 used a custom visualization inspired by *"A night under the stars A look at overnight stays at US National Parks"* [42], adapted for two fictional national parks: Crait and Jellystone, and altered the data to challenge typical U.S. seasonal temperature associations (e.g., winter is cold). For the experimental task, participants had three minutes to view a video or a static version of the visualization (Figure 2A-B).

The same visualization and caption (Appendix Table 3) are used in both conditions. However, there are three differences: (1) in the video condition, captions are narrated using Microsoft Azure AI Speech [1], with the corresponding text displayed on the screen (see Figures 2.1A and 2.1B), (2) during the narration, elements in the visualization are highlighted by dimming non-relevant visual elements (e.g., when the word "lodging" is read, all visuals except for the gold ring are dimmed); (3) participants can use a slider to revisit parts of the video. The total length of the video is 58 seconds.

After viewing the stimuli, participants were shown the same bespoke visualization, but for Jellystone National Park (Figure 2.2), and completed a 20-question multiple choice questionnaire (Appendix Table 4) designed to measure their comprehension. We had the first half of the questions focused on the visual encodings present in the visualization and the remaining focus on testing the participant's ability to translate what they know into predictions or trip planning.

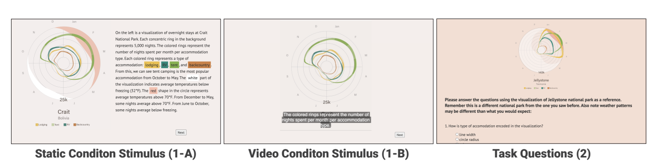

Figure 2: Screenshots of the central task. (1-A) Stimulus for static condition. (1-B) Stimulus for video condition. (2) Reference image for Jellystone National Park on the multiple-choice test post-stimulus.

4.0.3 Measurements. We collected the following from participants: demographic data, media and technology exposure questions, time spent on stimuli, responses to task questions, time to answer questions, engagement question answers, NASA TLX question answers, Mini-VLAT question answers, and time to complete the experiment.

4.0.4 Analysis and Results. Participants were excluded from analysis if they were colorblind, took unusually short time to complete the study (more than 3 standard deviations below mean), gave low-effort answers, did not finish the experiment, or did not play the video at least once (in the video condition). After data cleaning, 32 participants remained in the static condition, and 34 remained in the video condition. Participant demographics are shown in Appendix Table 1. For each remaining participant, performance on each comprehension question was graded, and accuracy was calculated as a sum across all questions.

Accuracy: We found no statistically significant differences in the task scores. A Shapiro-Wilk test for normality showed accuracy to be normally distributed in both conditions ($W = 0.971, p = 0.482$ (static), $W = 0.959, p = 0.22$ (video)). We ran a two-sample t-test for $accuracy\ stimuli$ and found no significant difference in accuracy across data videos and static visualization ($t(67) = 0.05, p = 0.96$). As a result, we fail to reject $\mathbf{H_0(1A)}$, suggesting no difference in visual encoding comprehension between data videos and static visualizations. Results are summarized in Figure 3.

Visualization Viewing Time: We found no statistically significant difference between the participants' viewing times for static and video. As shown in Figure 4, the data for the video condition exhibits a bimodal distribution, so we employed the Two-sample Kolmogorov–Smirnov test to compare the two distributions. The test results ($D = 0.261, p = 0.169$) indicate that we fail to reject the null hypothesis $\mathbf{H_0(1D)}$.

Subjective Ratings: We found statistically significant differences in two of the questions in the engagement questionnaire between conditions (Figure 5): "The visualization was fun to explore" ($\chi^2(5, N = 66) = 15.064, p = 0.01$), and "I feel like I could construct a story based on what I learned" ($\chi^2(5, N = 66) = 14.373, p = 0.013$). Due to the categorical nature of the data, we use the Chi-Square test for each of the questions.

We found no statistically significant differences in the NASA-TLX results between the two conditions (see Figure 6). For each question, we conducted a Shapiro-Wilk test to assess normality. We performed Mann-Whitney U tests as either one or both of the distributions were not normal. While we fail to reject the null hypothesis $\mathbf{H_0(1B)}$ in the NASA-TLX questionnaires. However, we find evidence to reject the null hypothesis regarding engagement.

Visual Literacy: We found no statistically significant difference between the visual literacy requirements for the two conditions (see relationship between participants' visual literacy scores (VLAT) and their task performance in Appendix Figure 8 and 9). We use analysis of covariance (ANCOVA) to compare regression models, with the task score as the dependent variable, the VLAT score as the covariate, and the condition (static visualization vs. data video) as the factor. The results ($F(1, 63) = 0.881, p = 0.352$) indicate non-rejection of the null hypothesis $\mathbf{H_0(1C)}$. We calculated Pearson's r and found moderate correlations between the two variables (VLAT Task Score) [2]: static condition ($r(30) = 0.46, p < 0.01$), data video condition ($r(32) = 0.458, p < 0.01$). The correlations are similar across the two conditions, so we cannot reject $\mathbf{H_0(1C)}$.

4.0.5 Discussion. We found no significant differences between condition subjective scores (aside from two engagement questions, discussed further in Section 6), visualization literacy requirements, or timing between the two conditions. Though we found no statistically significant difference in task scores between the two conditions in Experiment 1, we noticed a slight difference in mean and median scores.

5 Experiment 2: Complex and Unfamiliar Visualization

Results from Experiment 1 suggested that spatial references may have an impact on a person's understanding of data, but a potential ceiling effect may be present. In Experiment 2, we replaced the unfamiliar, but relatively simple national park visualization with a parallel coordinates plot (PCP) - a representation that is more visually complex, contains more data dimensions, and may benefit more from spatial references. The challenges associated with interpreting PCPs are well-documented in the visualization literature, both in interpreting a data-point as a poly-line, and the difficulty in recognizing the visual patterns of correlations (e.g. an 'X' shape in figure 7 indicates a negative correlation) [14].

By using a more complex visualization and focusing on spatial referencing, we hypothesize a more pronounced difference between conditions than in Experiment 1. In Experiment 2 we test the hypotheses: Given a complex, new visualization design there is no difference in: $\mathbf{H_0(2A)}$: data comprehension, $\mathbf{H_0(2B)}$: cognitive effort, $\mathbf{H_0(2C)}$: required visual literacy, and $\mathbf{H_0(2D)}$: visualization viewing times between participants.

5.0.1 Set-Up. The experimental setup was identical to that of Experiment 1, with the exception that the stimuli changed (see below), as did the set of questions asked to assess comprehension (Appendix Table 5). Experiment 2 was performed in two runs. In both runs, participants were asked post-hoc if they were familiar with parallel coordinates. As the retention of participants was below 50% in the first run, we introduced a screener question where participants were barred from participating if they indicated familiarity with parallel coordinates.

5.0.2 Stimuli. We "stacked the deck" in favor of data videos in Experiment 2. We incorporated explanatory visual cues in the data video to help identify complex patterns within the parallel coordinates. Figure 7 shows the three highlighted patterns in the video; outlier, positive relationship, and negative relationship. The duration of the video was 1 minute and 42 seconds long, and we gave participants 5 minutes and 20 seconds to view the stimulus. The static and video conditions share the same narrative (Appendix Table 3).

5.0.3 Analysis and Results. Participants with prior experience with parallel coordinates were dropped. The calculation of accuracy was based on a rubric developed by the authors (Appendix Table 5). After data cleaning, 93 participants remained in the static condition,

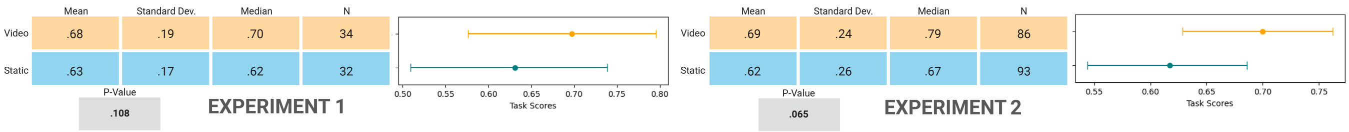

Figure 3: (Left) Descriptive statistics for each experiment with (Right) chart showing Mean and 95% confidence interval.

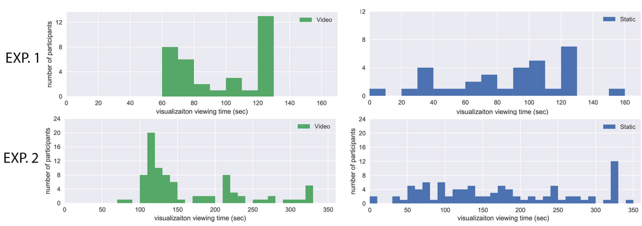

Figure 4: Time with stimulus for both experiments and conditions in seconds.

and 86 remained in the video condition. Participant demographics are shown in Appendix Table 2.

Accuracy: We found no significant differences in the task scores between the two conditions. A Shapiro-Wilk test showed accuracy to be non-normally distributed in both conditions ($W = 0.94, p = 0.0006$ (static); $W = 0.9, p = 0.0015$ (video)). We ran a two-tailed Mann-Whitney test and found no significant difference in accuracy across data videos and static visualization ($U = 3361.5, p = 0.065$). We fail to reject $\mathbf{H_0(2A)}$, *suggesting there is no difference in data comprehension between data videos and static visualizations, even if participants are unfamiliar with the visualization and the data video contains more information than the static visualization.* Results are summarized in Figure 3.

Visualization Viewing Time: We found no significant difference between participant viewing times for static visualizations and data videos (Figure 4). We use the Two-sample Kolmogorov–Smirnov test to compare the two distributions. The test results ($D = 0.277, p = 0.002$) indicate that we reject the null hypothesis $\mathbf{H_0(2E)}$.

Subjective Rating: For each question, we conducted a Shapiro-Wilk test to assess normality. We performed Mann-Whitney U tests as either one or both of the distributions were not normal. We found two statistically significant dimensions of the NASA-TLX test: Mental Demand ($U = 4705, p = 0.039$), Effort ($U = 4672.5, p = 0.049$). As shown in Figure 6, participants reported higher mental demand and higher effort for static visualizations.

We found no significant differences in the engagement questionnaire between the two conditions (Figure 5). For each question, the results are as follows: "I found the content easy to understand" ($\chi^2(5, N = 179) = 0.1972, p = 0.999$), "I found my mind wandering during these tasks" ($\chi^2(5, N = 179) = 0.0276, p = 1.000$), "The visualization was fun to explore" ($\chi^2(5, N = 179) = 0.1048, p = 0.999$), and "I feel like I could construct a story based on what I learned" ($\chi^2(5, N = 179) = 0.0071, p = 1.000$). While we fail to reject the null hypothesis in the engagement questionnaires, we find evidence to reject the null hypothesis regarding the Mental Demand and the Effort questions in the NASA-TLX.

Visual Literacy: We found no statistically significant difference between the visual literacy (Appendix Figure 10 and 11). We use analysis of covariance (ANCOVA) to compare the two regression models with the same parameters from Experiment 1. The test results ($F(1, 176) = 2.53, p = 0.114$) indicate that we fail to reject the null hypothesis $\mathbf{H_0(2D)}$.

5.0.4 Discussion. Our results show no statistically significant difference in task scores. However, participants reported significantly higher mental demand and effort for static visualizations on the NASA-TLX scale, suggesting more cognitive effort is required compared to data videos. There were no significant differences in other NASA-TLX measures or engagement metrics, indicating similar perceived understanding across both conditions. Visual literacy requirements were also similar, consistent with Experiment 1. We found a significant difference in viewing time: participants watched the data video once or twice, while static groups spent more time viewing the visualization and reading the caption. This aligns with the higher mental demand and effort scores for static visualizations.

Our findings suggest that data videos do not significantly outperform static visualizations, even for complex visualizations. This challenges the assumption that data videos' affordances, such as just-in-time explanations and synchronized highlighting, offer substantial advantages. We reflect on these results and their implications in the next section.

6 Discussion of Results, Limitations, and Future Work

In this section, we discuss the implications of our findings, their limitations, and future directions worth investigating.

What can a comparison of two single visualizations tell us about data videos? Running a controlled study required us to narrow the design and task space, limiting the generalizability of our results to effects of the specific category of data video (C3). Controlling the information shared across data videos and static visualizations to isolate the medium's effects (e.g., spoken narrative mirroring written text) may not reflect the richness of data videos in the wild, which often include narrative tone, music, and additional media. Additionally, our study focused on data videos presenting a single visualization at a time, while many data videos use multiple visualizations to highlight various data aspects. We acknowledge further exploration is needed to definitively identify affordances and limitations of data videos, but believe investigating these more complex examples requires first understanding their foundational building blocks.

Why care about a null result? Although null results are often overlooked in academic publishing, our finding is noteworthy in its paradoxical nature – given the popularity and preference for data videos versus their actual benefits in comprehension and cognitive

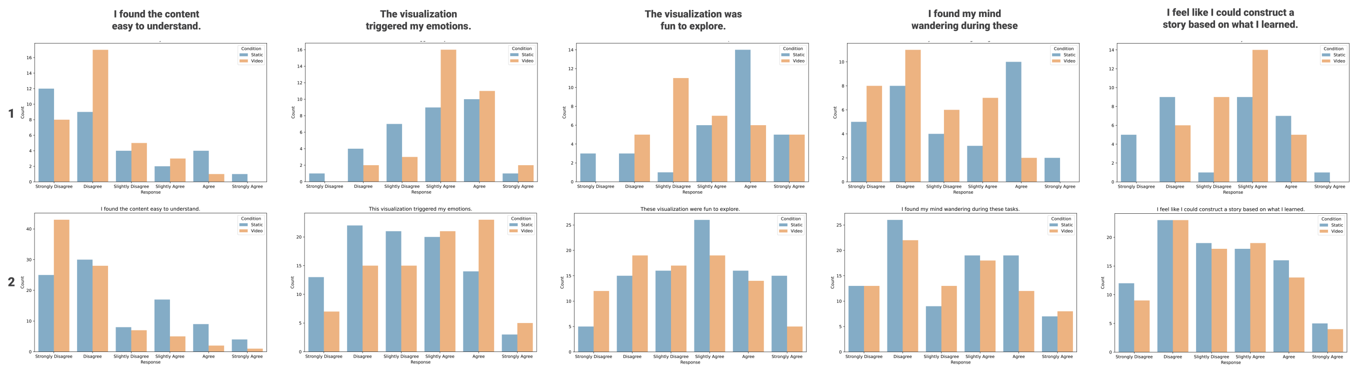

Figure 5: Answers for the post-task engagement questions from our experiments. For all sub-figures, the static condition (blue) is represented as 0 and the video (orange) is represented as 1.

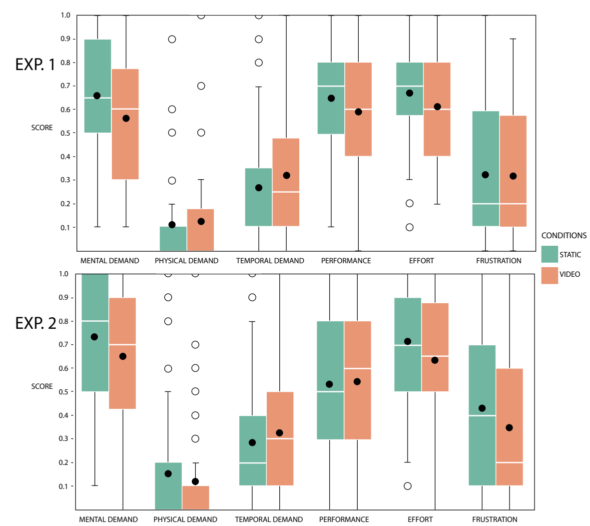

Figure 6: NASA-TLX Scores for Experiment 1 (Top) and Experiment 2 (Bottom).

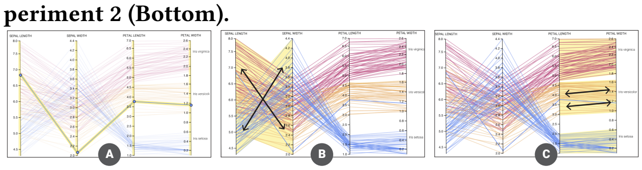

Figure 7: Visual Cues highlighting patterns in the parallel coordinates plot. (A) Outlier highlighting (B) Negative relationship Indicated by Crossed Lines. (C) Positive relationship indicated by parallel lines.

workload. Recent work suggests examining data distributions instead of relying solely on significance testing. In Figure 3, the distributions appear distinguishable for Experiments 1 and 2, indicating potential effectiveness for certain participant groups. Identifying commonalities among user subgroups benefiting from data videos remains an open challenge. While we do not draw strong conclusions about the positive effects of data videos for the population at large, we also do not close the door on its value to the community.

Do people actually prefer data videos? Previous studies indicate that participants prefer data videos to static visualizations [4], but the results of our engagement questionnaire do not align with these findings. We found no significant differences between conditions for our engagement questions, except for two questions in Experiment 1: *"The visualization was fun to explore,"* and *"I feel like I could construct a story based on what I learned"*. Further, it appears the distributions for static are generally more in agreement that the *"The visualization was fun to explore."*

We do not interpret these results as evidence that people do not prefer data videos, but that data videos might not be a clear preference for everyone. For example, it might depend on individual acuity to different types of learning [24]. We believe that to identify preference reliably, further work should be done within these categorizations of participants.

Where do we go from here? Null results in academia can be seen as "dead ends," but we do not believe that is the case here. The widespread use and sharing of data videos on social platforms suggest that researchers should continue to study their impacts. Our experiments explore a small segment of a broader field of design and evaluation possibilities. We focused on the dynamic, temporal qualities of data videos, but the larger world of video design intersects with media expertise and disciplines rarely drawn upon in information visualization.

The integration of data videos with social media may require new evaluations from social computing. Questions such as how frequently data videos are shared compared to static visualizations, how sharing behavior propagates across social networks, and how conversations differ between formats demand different evaluation methods and tools than traditional visualization studies.

7 Conclusion

In our investigation into the efficacy of data videos compared to static visualizations for data comprehension, we found no significant differences across experiments. Despite hypothesizing that the dynamic and multi-sensory nature of data videos would benefit understanding, our results suggest that data videos are not inherently more effective than static visualizations. This outcome prompts a reassessment of the assumed benefits of data videos, highlighting the need for further research to explore the trade-offs between user engagement and the cognitive demands of these formats.

Future studies should investigate the contextual and situational factors that might influence the effectiveness of data videos. Nonetheless, these initial results can inform practitioners who are creating data videos for social or web platforms. to weigh the trade-offs between video and static formats—particularly regarding user comprehension and engagement.

References

[1] [n. d.]. Microsoft Azure AI Speech. https://azure.microsoft.com/products/ai-services/ai-speech/. Accessed: 2024-9-12.

[2] Haldun Akoglu. 2018. User's guide to correlation coefficients. *Turkish journal of emergency medicine* 18, 3 (2018), 91–93.

[3] Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Christophe Hurter, and Pourang Irani. 2015. Understanding data videos: Looking at narrative visualization through the cinematography lens. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1459–1468. https://doi.org/10.1145/2702123.2702431

[4] Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Jason Leboe-McGowan, and Pourang Irani. 2018. Hooked on data videos: assessing the effect of animation and pictographs on viewer engagement. In *Proceedings of the 2018 international conference on advanced visual interfaces*. 1–9.

[5] Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Andres Monroy-Hernandez, and Pourang Irani. 2017. Authoring Data-Driven Videos with DataClips. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 501–510. https://doi.org/10.1109/TVCG.2016.2598647

[6] Daniel Archambault, Helen Purchase, and Bruno Pinaud. 2011. Animation, Small Multiples, and the Effect of Mental Map Preservation in Dynamic Graphs. *IEEE Transactions on Visualization and Computer Graphics* 17, 4 (2011), 539–552. https://doi.org/10.1109/TVCG.2010.78

[7] Sandra Berney and Mireille Bétrancourt. 2016. Does animation enhance learning? A meta-analysis. *Computers and Education* 101 (2016), 150–167. https://doi.org/10.1016/j.compedu.2016.06.005

[8] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D3: Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (dec 2011), 2301–2309. https://doi.org/10.1109/TVCG.2011.185

[9] M. Brehmer, B. Lee, P. Isenberg, and E. Choe. 2020. A Comparative Evaluation of Animation and Small Multiples for Trend Visualization on Mobile Phones. *IEEE Transactions on Visualization amp; Computer Graphics* 26, 01 (jan 2020), 364–374. https://doi.org/10.1109/TVCG.2019.2934397

[10] Qing Chen, Shixiong Cao, Jiazhe Wang, and Nan Cao. 2023. How does automation shape the process of narrative visualization: A survey of tools. *IEEE Transactions on Visualization and Computer Graphics* (2023).

[11] Zhutian Chen, Qisen Yang, Jiarui Shan, Tica Lin, Johanna Beyer, Haijun Xia, and Hanspeter Pfister. 2023. iBall: Augmenting Basketball Videos with Gaze-moderated Embedded Visualizations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–18. https://doi.org/10.1145/3544548.3581266 arXiv:2303.03476

[12] Hao Cheng, Junhong Wang, Yun Wang, Bongshin Lee, Haidong Zhang, and Dongmei Zhang. 2022. Investigating the role and interplay of narrations and animations in data videos. In *Computer Graphics Forum*, Vol. 41. Wiley Online Library, 527–539.

[13] Fanny Chevalier, Nathalie Henry Riche, Catherine Plaisant, Amira Chalbi, and Christophe Hurter. 2016. Animations 25 years later: New roles and opportunities. In *Proceedings of the Workshop on Advanced Visual Interfaces, AVI'16*. 280–287. https://doi.org/10.1145/2909132.2909255

[14] Elias Elmquist, Kajetan Enge, Alexander Rind, Carlo Navarra, Robert Höldrich, Michael Iber, Alexander Bock, Anders Ynnerman, Wolfgang Aigner, and Niklas Rönnberg. 2024. Parallel Chords: an audio-visual analytics design for parallel coordinates. *Personal and Ubiquitous Computing* (may 2024). https://doi.org/10.1007/s00779-024-01795-8

[15] Tong Ge, Yue Zhao, Bongshin Lee, Donghao Ren, Baoquan Chen, and Yunhai Wang. 2020. Canis: A High-Level Language for Data-Driven Chart Animations. *Computer Graphics Forum* 39, 3 (2020), 607–617. https://doi.org/10.1111/cgf.14005

[16] Brian D. Hall, Lyn Bartram, and Matthew Brehmer. 2022. Augmented Chironomia for Presenting Data to Remote Audiences. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology, UIST'22*. ACM, New York, NY, USA, 1–14. https://doi.org/10.1145/3526113.3545614

[17] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[18] Younghoon Kim and Jeffrey Heer. 2021. Gemini: A Grammar and Recommender System for Animated Transitions in Statistical Graphics. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (feb 2021), 485–494. https://doi.org/10.1109/TVCG.2020.3030360 arXiv:2009.01429

[19] Younghoon Kim and Jeffrey Heer. 2021. Gemini2: Generating Keyframe-Oriented Animated Transitions between Statistical Graphics. In *Proceedings of the 2021 IEEE Visualization Conference, VIS'21 Short Paper*. 201–205. https://doi.org/10.1109/VIS49827.2021.9623291 arXiv:2108.04385

[20] Ha-Kyung Kong, Wenjie Zhu, Zhicheng Liu, and Karrie Karahalios. 2019. Understanding Visual Cues in Visualizations Accompanied by Audio Narrations. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300280

[21] Tica Lin, Zhutian Chen, Yalong Yang, Daniele Chiappalupi, Johanna Beyer, and Hanspeter Pfister. 2023. The Quest for Omnioculars: Embedded Visualization for Augmenting Basketball Game Viewing Experiences. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2023), 962–971. https://doi.org/10.1109/TVCG.2022.3209353

[22] R. K. Lowe. 2003. Animation and learning: Selective processing of information in dynamic graphics. *Learning and Instruction* 13, 2 (2003), 157–176. https://doi.org/10.1016/S0959-4752(02)00018-X

[23] Min Lu, Noa Fish, Shuaiqi Wang, Joel Lanir, Daniel Cohen-Or, and Hui Huang. 2022. Enhancing Static Charts With Data-Driven Animations. *IEEE Transactions on Visualization and Computer Graphics* 28, 7 (jul 2022), 2628–2640. https://doi.org/10.1109/TVCG.2020.3037300

[24] Robert J Murphy, Sarah A Gray, Sorin R Straja, and Meredith C Bogert. 2004. Student learning preferences and teaching implications. *Journal of dental education* 68, 8 (2004), 859–866.

[25] Nic Newman, Richard Fletcher, Craig T Robertson, Amy Ross Arguedas, and Rasmus Kleis Nielsen. 2024. Reuters Institute digital news report 2024. *Reuters Institute for the study of Journalism* (2024).

[26] Humphrey O. Obie, Caslon Chua, Iman Avazpour, Mohamed Abdelrazek, John Grundy, and Tomasz Bednarz. 2019. A study of the effects of narration on comprehension and memorability of visualisations. *Journal of Computer Languages* 52, April (jun 2019), 113–124. https://doi.org/10.1016/j.cola.2019.04.006

[27] Humphrey O. Obie, Caslon Chua, Iman Avazpour, Mohamed Abdelrazek, John Grundy, and Tomasz Bednarz. 2019. A study of the effects of narration on comprehension and memorability of visualisations. *Journal of Computer Languages* 52 (2019), 113–124. https://doi.org/10.1016/j.cola.2019.04.006

[28] Yang Ouyang, Leixian Shen, Yun Wang, and Quan Li. 2024. NotePlayer: Engaging Computational Notebooks for Dynamic Presentation of Analytical Processes. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. ACM, New York, NY, USA, 1–20. https://doi.org/10.1145/3654777.3676410

[29] Saugat Pandey and Alvitta Ottley. 2023. Mini-VLAT: A Short and Effective Measure of Visualization Literacy. In *Computer Graphics Forum*, Vol. 42. Wiley Online Library, 1–11.

[30] Prolific. 2014. *Prolific Online Platform*. Retrieved August 9, 2024 from https://www.prolific.com

[31] George Robertson, Roland Fernandez, Danyel Fisher, Bongshin Lee, and John Stasko. 2008. Effectiveness of Animation in Trend Visualization. *IEEE Transactions on Visualization and Computer Graphics* 14, 6 (2008), 1325–1332. https://doi.org/10.1109/TVCG.2008.125

[32] Hans Rosling. [n. d.]. *The best stats you've ever seen*. https://www.ted.com/talks/hans_rosling_the_best_stats_you_ve_ever_seen?subtitle=en

[33] Sadia Rubab, Lingyun Yu, Junxiu Tang, and Yingcai Wu. 2023. Exploring Effective Relationships Between Visual-Audio Channels in Data Visualization. *Journal of Visualization* 26, 4 (2023), 937–956. https://doi.org/10.1007/s12650-023-00909-3

[34] Edward Segel and Jeffrey Heer. 2010. Narrative visualization: Telling stories with data. *IEEE transactions on visualization and computer graphics* 16, 6 (2010), 1139–1148.

[35] Zekai Shao, Leixian Shen, Haotian Li, Yi Shan, Huamin Qu, Yun Wang, and Siming Chen. 2025. Narrative Player: Reviving Data Narratives with Visuals. *IEEE Transactions on Visualization and Computer Graphics* (2025), 1–15. https://doi.org/10.1109/TVCG.2025.3530512

[36] Leixian Shen, Haotian Li, Yun Wang, Tianqi Luo, Yuyu Luo, and Huamin Qu. 2024. Data Playwright: Authoring Data Videos With Annotated Narration. *IEEE Transactions on Visualization and Computer Graphics* (2024), 1–14. https://doi.org/10.1109/TVCG.2024.3477926

[37] Leixian Shen, Haotian Li, Yun Wang, and Huamin Qu. 2024. From Data to Story: Towards Automatic Animated Data Video Creation with LLM-Based Multi-Agent Systems. In *IEEE VIS 2024 Workshop on Data Storytelling in an Era of Generative AI, GEN4DS'24*. IEEE, 20–27. https://doi.org/10.1109/GEN4DS63889.2024.00008

[38] Leixian Shen, Haotian Li, Yun Wang, and Huamin Qu. 2025. Reflecting on Design Paradigms of Animated Data Video Tools. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, 1–21. https://doi.org/10.1145/3706598.3713449

[39] Leixian Shen, Yizhi Zhang, Haidong Zhang, and Yun Wang. 2024. Data Player: Automatic Generation of Data Videos with Narration-Animation Interplay. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2024), 109–119. https://doi.org/10.1109/TVCG.2023.3327197

[40] Wai Tong, Kento Shigyo, Lin-ping Yuan, Mingming Fan, Ting-chuen Pong, Huamin Qu, and Meng Xia. 2024. VisTellAR: Embedding Data Visualization to Short-form Videos Using Mobile Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics* (2024), 1–13. https://doi.org/10.1109/TVCG.2024.3372104

[41] Barbara Tversky, Julie Bauer Morrison, and Mireille Betrancourt. 2002. Animation: can it facilitate? *International Journal of Human-Computer Studies* 57, 4 (2002), 247–262. https://doi.org/10.1006/ijhc.2002.1017

[42] Jordan Vincent. 2018. *A night under the stars A look at overnight stays at US National Parks.* Retrieved August 9, 2024 from https://jordan-vincent.com/night-under-the-stars/

[43] Vox. [n. d.]. How tax brackets actually work. https://www.youtube.com/watch?v=VJhsjUPDulw&list=PLJ8cMiYb3G5cOIKx13ZmhoruAAF83lP74&index=9&ab_channel=Vox

[44] Yun Wang, Leixian Shen, Zhengxin You, Xinhuan Shu, Bongshin Lee, John Thompson, Haidong Zhang, and Dongmei Zhang. 2024. WonderFlow: Narration-Centric Design of Animated Data Videos. *IEEE Transactions on Visualization and Computer Graphics* (2024), 1–17. https://doi.org/10.1109/TVCG.2024.3411575

[45] Lijie Yao, Romain Vuillemot, Anastasia Bezerianos, and Petra Isenberg. 2024. Designing for Visualization in Motion: Embedding Visualizations in Swimming Videos. *IEEE Transactions on Visualization and Computer Graphics* 30, 3 (2024), 1821–1836. https://doi.org/10.1109/TVCG.2023.3341990

[46] Lu Ying, Yun Wang, Haotian Li, Shuguang Dou, Haidong Zhang, Xinyang Jiang, Huamin Qu, and Yingcai Wu. 2024. Reviving static charts into live charts. *IEEE Transactions on Visualization and Computer Graphics* (2024).

[47] Tongyu Zhou, Jeff Huang, and Gromit Yeuk-yin Chan. 2024. Epigraphics: Message-Driven Infographics Authoring. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–18. https://doi.org/10.1145/3613904.3642172

Appendix Table 1:		
Experiment 1: Static N = 32, Video N = 34		
Measure	**Static Group Percentages**	**Video Group Percentages**
Age	18-24: **35%**, 25-39: **34%**, 40-49: **9%**, 50-59: **0**, 60+: **0**	18-24: **44%**, 25-39: **47%**, 40-49: **0**, 50-59: **9%**, 60+: **0**
Degree	High School: **43%**, Bachelors: **34%**, Masters: **12%**, PhD: **3%**, Other: **6%**, Prefer Not To Answer: **0**	High School: **41%**, Bachelors: **35%**, Masters: **23%**, PhD: **0**, Other: **0**, Prefer Not To Answer: **0**
About how many hours per week do you use a computer?	lt5: **0**, 6-10: **3%**, 11-20: **9%**, 21-30: **31%**, 31-40: **9%**, 41-50: **6%**, 51-60: **12%**, gt60: **28%**	lt5: **3%**, 6-10: **3%**, 11-20: **12%**, 21-30: **23%**, 31-40: **11%**, 41-50: **14%**, 51-60: **14%**, gt60: **17%**
Which of the following statements best describes your experience with data visualizations?	No experience: **3%**, Viewed but not created: **53%**, Modified existing: **15%**, Created from scratch: **28%**, Prefer not answer: **0**	No experience: **3%**, Viewed but not created: **44%**, Modified existing: **29%**, Created from scratch: **23%**, Prefer not answer: **0**
How often do you use data visualization and/or analysis tools (e.g., Excel, Tableau)?	Never: **12%**, Rarely: **28%**, Occasionally: **28%**, Frequently: **29%**	Never: **3%**, Rarely: **44%**, Occasionally: **23%**, Frequently: **0**
Considering your exposure to digital media, how often have you engaged with data visualizations?	Never/Rarely: **0**, Sometimes: **34%**, Regularly: **18%**, Frequently: **47%**, Prefer not to answer: **0**	Never/Rarely: **6%**, Sometimes: **32%**, Regularly: **32%**, Frequently: **0**, Prefer not to answer: **0**
How do you approach learning with complex digital information?	Skip: **0**, Seek summaries: **28%**, Research details: **40%**, Analyze thoroughly: **28%**, Prefer not answer: **3%**	Skip: **3%**, Seek summaries: **38%**, Research details: **28%**, Analyze thoroughly: **20%**, Prefer not answer: **0**
How often do you engage with digital content (e.g., articles, reports, videos)?	Rarely: **0**, Occasionally: **12%**, Often: **44%**, Daily: **44%**, Prefer not answer: **0**	Rarely: **3%**, Occasionally: **23%**, Often: **35%**, Daily: **38%**, Prefer not answer: **0**

Table 1: Participant Answers to Demographic and Media Exposure Questions

Appendix Table 2:		
Experiment 2: Static N = 34, Video N = 35		
Measure	**Static Group Percentages**	**Video Group Percentages**
Age	18-24: **43%**, 25-39: **51%**, 40-49: **3%**, 50-59: **0**, 60+: **3%**	18-24: **37%**, 25-39: **41%**, 40-49: **0**, 50-59: **8%**, 60+: **3%**
Degree	High School: **28%**, Bachelors: **57%**, Masters: **12%**, PhD: **0**, Other: **0**, Prefer Not To Answer: **3%**	High School: **23%**, Bachelors: **54%**, Masters: **11%**, PhD: **3%**, Other: **6%**, Prefer Not To Answer: **3%**
About how many hours per week do you use a computer?	lt5: **0**, 6-10: **17%**, 11-20: **9%**, 21-30: **6%**, 31-40: **34%**, 41-50: **11%**, 51-60: **20%**, gt60: **8%**, Prefer Not To Answer: **3%**	lt5: **6%**, 6-10: **6%**, 11-20: **3%**, 21-30: **17%**, 31-40: **14%**, 41-50: **14%**, 51-60: **14%**, gt60: **20%**, Prefer Not To Answer: **6%**
Which of the following statements best describes your experience with data visualizations?	No experience: **8%**, Viewed but not created: **51%**, Modified existing: **11%**, Created from scratch: **28%**, Prefer not answer: **0**	No experience: **5%**, Viewed but not created: **51%**, Modified existing: **14%**, Created from scratch: **25%**, Prefer not answer: **3%**
How often do you use data visualization and/or analysis tools (e.g., Excel, Tableau)?	Never: **6%**, Rarely: **40%**, Occasionally: **14%**, Frequently: **40%**, Prefer not to answer: **0**	Never: **5%**, Rarely: **14%**, Occasionally: **23%**, Frequently: **54%**, Prefer not to answer: **3%**
Considering your exposure to digital media, how often have you engaged with data visualizations?	Never/Rarely: **8%**, Sometimes: **34%**, Regularly: **17%**, Frequently: **40%**, Prefer not to answer: **0**	Never/Rarely: **6%**, Sometimes: **26%**, Regularly: **31%**, Frequently: **34%**, Prefer not to answer: **3%**
How do you approach learning with complex digital information?	Skip: **3%**, Seek summaries: **31%**, Research details: **40%**, Analyze thoroughly: **26%**, Prefer not answer: **0**	Skip: **0%**, Seek summaries: **37%**, Research details: **26%**, Analyze thoroughly: **37%**, Prefer not answer: **0**
How often do you engage with digital content (e.g., articles, reports, videos)?	Rarely: **3%**, Occasionally: **28%**, Often: **31%**, Daily: **37%**, Prefer not answer: **0**	Rarely: **3%**, Occasionally: **17%**, Often: **46%**, Daily: **31%**, Prefer not answer: **0**

Table 2: Participant Answers to Demographic and Media Exposure Questions (Experiment 2)

Appendix Table 3:	
Experiment 1-2: Caption/Narration	
Experiment 1	**Experiment 2**
On the left is a visualization of overnight stays at Crait National Park. Each concentric ring in the background represents 5,000 nights. The colored rings represent the number of nights spent per month per accommodation type. Each colored ring represents a type of accommodation: lodging, RV, tent, and backcountry. *From this, we can see tent camping is the most popular accommodation from October to May. The white part of the visualization indicates average temperatures below freezing (32°F). The red shape in the circle represents average temperatures above 70°F.* *From December to May, some nights average above 70°F. From June to October, some nights average below freezing.*	*This parallel coordinates plot presents a multivariate analysis of the classic Iris dataset. The dataset has four numerical variables, shown as four distinct axes in the plot: sepal length, sepal width, petal length, and petal width.* *A single polyline represents an individual data point across multiple variables. The values for a given data point are represented as the intersection of the polyline at each axis. The polylines in this plot are color-coded by three iris species: the blue lines for Iris setosa, the magenta lines for Iris virginica, and the orange lines for Iris versicolor.* *Parallel coordinates enable us to explore the relative differences and similarities among the three Iris species across the four variables in the dataset. The plot reveals relationships between the variables across all data points.* *For example, there is a noticeable negative relationship between Sepal Length and Sepal Width, as indicated by the intersecting "X" pattern of the lines. This pattern shows that when Sepal Length increases, Sepal Width generally decreases. In contrast, Petal Length and Petal Width demonstrate a positive relationship through the parallel arrangement of the lines, meaning that as one variable increases, the other tends to increase as well.*

Table 3: Caption and narration for Experiment 1 and Experiment 2 visualizations.

Appendix Table 4: Experiment 1-2: Post-Stimuli Multiple-Choice Questions	
1. How is type of accommodation encoded in the visualization?	2. What does the central number "140k" most likely represent?
3. The chart most likely represents data collected over what time frame?	4. What does the red space in the visualization indicate about the average temperature?
5. What does each colored line represent?	6. When a colored line gets closer to the edge of the chart, what does this mean?
7. What accommodation is used the least year-round?	8. Which month appears to have the highest overall visitation regardless of accommodation type?
9. What does the outermost ring of the chart likely indicate?	10. How many different categories of accommodation are there?
11. Your friend only wants to visit when it's likely to be between 32°F-70°F. Which month will you recommend?	12. What is the nightly peak of tent camping activity in Jellystone?
13. What type of accommodation is available all year round?	14. You plan to spend time in Jellystone to snowshoe. Please select all accommodations available at that time of year in Jellystone.
15. You are planning a trip to Jellystone but want to choose the least popular accommodation to beat the crowds. What type of accommodation has the least amount of nights spent per year?	16. You are making recommendations for a colleague's trip to Jellystone. They're sensitive to temperature and prefer visiting the park when the average temperature is above freezing but below 70°F. What month is in the ideal timeframe to recommend?
17. Which statement is true regarding the popularity of accommodations in Jellystone Park?	18. What might be a reason for the small bump in lodging during August, as indicated in the visualization?
19. Based on the seasonal patterns shown in the visualization, if you wanted to avoid both extreme temperatures and high visitor numbers for a peaceful trip, which month might be the optimal choice?	20. Given the trends shown in the visualization, if the park introduced a new winter festival in July to celebrate mid-winter, how might this event be reflected in the data for lodging?

Table 4: Post-Stimuli Multiple-Choice Questions for Experiment 1-2

Appendix 5:	
Experiment 1 Post-Stimuli Open-Text Questions and Rubric Used for Grading	
1. What is the relationship between the values for Sepal Length and Petal Length?	**Correct Answer:** Positive Relationship. **Rubric:** 1 point: Answers that indicate a positive relationship using terms like "directly proportional" or "high in one means high in the other." 0.5 points: Answers that do not indicate a positive relationship or use incorrect terms such as "linear," "correlated," or "proportional" without specifying the type. 0 points: Answers like "the same" or any incorrect interpretation.
2. What is the relationship between the values for Sepal Width and Petal Length?	**Correct Answer:** Negative Relationship. **Rubric:** 1 point: Answers that indicate a negative relationship using terms like "inversely proportional," "high in one means low in the other," or "inverse." 0.5 points: Answers that do not indicate a negative relationship or use incorrect terms such as "linear," "correlated," or "proportional" without specifying the type. 0 points: Incorrect responses or unclear explanations.
3. For any iris flower to have a high sepal length, do they typically have a high, mid, or low Petal Width?	**Correct Answer:** High Petal Width **Rubric:** 1 point: Answers that correctly indicate "high" or "high to mid." 0 points: Answers that indicate "mid" alone, "low," or any incorrect relationship.
4. Reading the plot, if an Iris Setosa flower has a 0.4 Petal Width, what is the highest possible Sepal Length the flower could have?	**Correct Answer:** Greater than or equal to 5, less than 6 **Rubric:** 1 point: Answers that provide a number within the range of 5 to 5.9 or explicitly state this range. 0 points: Answers outside the correct range or that indicate an incorrect understanding of the plot.
5. Which attribute shows the densest clustering of lines, indicating high similarity in measurements within a species?	**Correct Answer:** Petal Width. **Rubric:** 1 point: Answers that correctly indicate "Petal Width." 0 points: Answers that indicate attributes other than "Petal Width."
6. Which species of Iris, if any, shows an outlier? Describe what makes it an outlier.	**Correct Answer:** Iris Setosa. **Rubric:** 1 point: Answers that correctly identify "Iris Setosa" and describe that one data point trends opposite to the other points' feature values. 0.5 points: Answers that correctly identify "Iris Setosa" but do not provide an explanation. 0 points: Answers that identify the wrong species or fail to describe what makes the identified species an outlier.

Table 5: Post-Stimuli Open-Text Questions and Grading Rubric for Experiment 1.

Appendix 6:

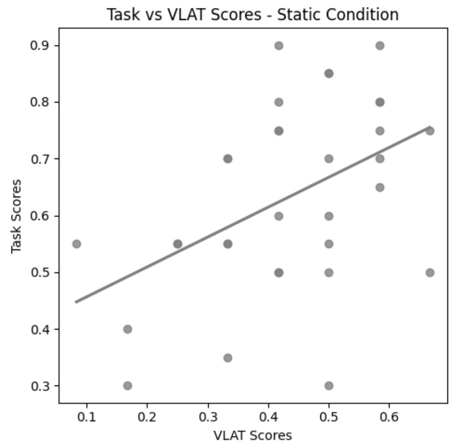

Figure 8: Experiment 1 Static Task vs. VLAT scores. Correlation: 0.460, p-value: 0.008

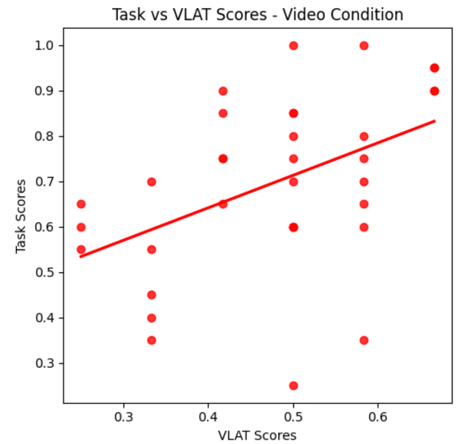

Figure 9: Experiment 1 Video Task vs. VLAT scores. Correlation: 0.458, p-value: 0.006

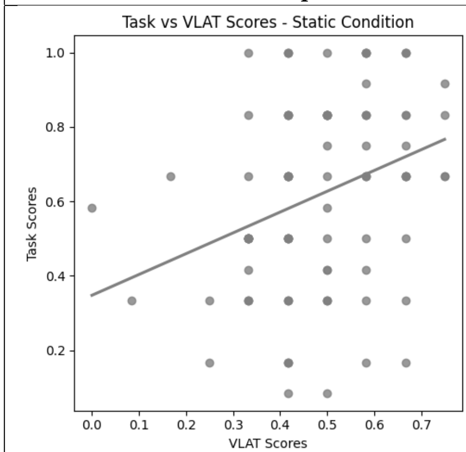

Figure 10: Experiment 2 Static Task vs. VLAT scores. Correlation: 0.315, p-value: 0.002

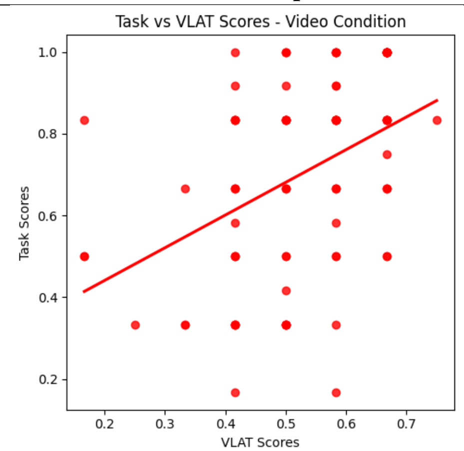

Figure 11: Experiment 2 Video Task vs. VLAT scores. Correlation: 0.411, p-value: 0.000

Table 6: VLAT vs Task Scores for Experiments 1-2