SHORT-PAPER

Enhancing Cognitive Clarity through Drill-Down Structuring in Data Videos

YONGQING CHEN, University of Technology Sydney, Sydney, NSW, Australia

JIE LIANG, University of Technology Sydney, Sydney, NSW, Australia

KAYE CHAN, University of Technology Sydney, Sydney, NSW, Australia

NINA ERREY, University of Technology Sydney, Sydney, NSW, Australia

CHENXUAN ZHOU, University of Technology Sydney, Sydney, NSW, Australia

YI CHEN, Beijing Technology and Business University, Beijing, Beijing, China

Open Access Support provided by:

University of Technology Sydney

Beijing Technology and Business University

Published: 01 December 2025

Citation in BibTeX format

VINCI 2025: Proceedings of the 18th International Symposium on Visual Information Communication and Interaction
December 1 - 3, 2025
Linz, Austria

VINCI '25: Proceedings of the 18th International Symposium on Visual Information Communication and Interaction (December 2025)
https://doi.org/10.1145/3769534.3769597
ISBN: 9798400718458

Enhancing Cognitive Clarity through Drill-Down Structuring in Data Videos

Yongqing Chen*
University of Technology Sydney
Sydney, Australia
yongqing.chen@student.uts.edu.au

Jie Liang
University of Technology Sydney
Sydney, Australia
Jie.Liang@uts.edu.au

Kaye Chan
University of Technology Sydney
Sydney, Australia
kaye.chan@uts.edu.au

Nina Errey
University of Technology Sydney
Sydney, Australia
nina.errey@uts.edu.au

Chenxuan Zhou
University of Technology Sydney
Sydney, China
chenxuanzhou90@gmail.com

Yi Chen
Beijing Technology and Business University
Beijing, China
chenyi@th.btbu.edu.cn

Abstract

Data videos are widely used in media and education, but can overwhelm viewers if poorly organized. We assess whether a hierarchical drill-down structure improves comprehension and reduces extraneous cognitive load in linear, non-interactive data videos. Building on cognitive load theory and narrative visualization research, we propose a conceptual model that divides a narrative into successive layers of detail. We conducted an online between-subjects experiment (N = 100) comparing a drill-down video with an equivalent flat baseline. To isolate visual-structuring effects and reflect common sound-off contexts (e.g., autoplay feeds, public displays), we used short, caption-free videos without audio. Independent-samples t-tests showed slightly better recall with drill-down but no statistically significant differences in recall, cognitive load, or self-reported comprehension. Qualitative feedback highlighted that fast pacing and high visual density in both videos imposed substantial cognitive demands, likely overshadowing any structural benefits. Our findings encourage designs that combine drill-down structuring with adaptive pacing, persistent visual anchors, and multimedia cues.

CCS Concepts

• **Human-centered computing → Empirical studies in HCI**.

Keywords

Narrative visualization, Data video, Drill-Down structuring, Cognitive load, User comprehension.

ACM Reference Format:
Yongqing Chen, Jie Liang, Kaye Chan, Nina Errey, Chenxuan Zhou, and Yi Chen. 2025. Enhancing Cognitive Clarity through Drill-Down Structuring in Data Videos. In *Proceedings of the 18th International Symposium on Visual Information Communication and Interaction (VINCI 2025), December 01–03, 2025, Linz, Austria.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3769534.3769597

*Corresponding author.

ACM ISBN 979-8-4007-1845-8/25/12
https://doi.org/10.1145/3769534.3769597

1 Introduction

Data videos are a powerful form of narrative visualization that integrate animated data graphics with storytelling [2, 3, 20]. They are used across journalism, education, marketing, and public campaigns [3, 20]. Despite their growing popularity, the definition of a "data video" varies inconsistently in previous works. Prior work uses the term variously for author-driven narratives with a strict linear path and no interactivity [20]; for data videos as motion graphics combining sound and visuals to tell a data story [3, 27]; and for designs leveraging animation and pictographs to increase engagement and sometimes comprehension [3]. In this study, we focus on linear, non-interactive, sound-off videos that incorporate animated data visualizations into storytelling. This format is often viewed on social platforms as well as in sound-restricted public contexts (e.g., digital signage, transport hubs, and classrooms) [24]. These videos are designed for passive consumption, meaning viewers cannot control the pace or sequence of information. Thus, effective structuring is essential for communicating dense content. Interactive systems support exploration and user pacing that can mitigate complexity [13, 19, 20, 23]. Multimedia learning also shows that structuring, especially segmentation, reduces extraneous load [16]. By contrast, linear and non-interactive data videos have received comparatively little empirical study on delivering detail without overload through structural design alone. In such passive settings, viewers have little control over pace or order, so designers need to manage flow and segmentation carefully [6, 16, 26].

Therefore, it is crucial to examine whether the inherent structure of the content, through sequencing and layering, can serve as a compensatory mechanism when interactivity is lacking. To address this gap, we propose a drill-down structuring model for linear, non-interactive data videos. This approach leverages cognitive load theory, chunking principles, and multimedia learning guidelines to direct attention, gradually reveal complexity, and support understanding. In the empirical experiment stage, we developed two versions of data videos: one with a drill-down structure and one with a flat linear baseline, while keeping all other factors constant. We then measured factual recall, cognitive load, and perceived comprehension. This paper contributes to research on data storytelling and HCI in two ways. First, we present a conceptual model for hierarchical drill-down structuring in data videos and empirically compare it with a flat baseline. Second, the results indicate that fast pacing and high visual density likely dominated the viewing experience, motivating adaptive pacing and persistent visual anchors in future designs, informing practice.

2 Related Work

2.1 Data Videos in Narrative Visualization

Data videos are recognized as a narrative-visualization genre that integrates storytelling with animated, data-driven graphics, as characterized by Segel and Heer [20]. Building on this line, Amini describes data videos as motion graphics incorporating visualizations and narrative devices [2], and Zhao and Elmqvist survey media types for data-driven storytelling, explicitly including video [27]. Regarding communicative intent, Cao distinguishes five goal-driven genres—factual event, factual knowledge, hypothesis, persuasion, and prediction [8]. Structurally, some videos are strictly linear with designer-set sequencing [20], while others adopt modular/segment-based organization via partitioning and sequence [28]. Consequently, we consider three dimensions: structure (linear vs. modular), interactivity (author-driven/non-interactive vs. reader-driven/interactive), and purpose. Differentiating from prior works, this study focuses on non-interactive, linear videos aimed at explanation commonly seen in social media and public-facing communication.

2.2 Cognitive Load in Passive Multimedia

Cognitive Load Theory and the Cognitive Theory of Multimedia Learning distinguish intrinsic, extraneous, and germane load, and recommend reducing extraneous load through design [16, 25]. In instructor-paced or interactive settings, techniques such as segmentation, signalling, and worked examples have been shown to improve comprehension by reducing unnecessary processing demands [16, 25]. However, most evidence comes from contexts with narration or user control rather than non-interactive, sound-off data videos. Furthermore, recent investigations suggest that cognitive overload may not only stem from information density but also from how the information is temporally organized [15, 16]. Moreover, excessive video playback speed can raise cognitive load and impair learning video-based study [18]. These gaps motivate our test of a purely structural intervention—hierarchical drill-down in a non-interactive, linear format.

2.3 Hierarchical Exploration and Drill-Down Structuring

Following Segel and Heer, we use structuring to mean author-driven linear ordering of scenes and discrete segmentation of the narrative flow [20]. Following Elmqvist and Fekete, we use hierarchical structuring for multiscale organization across levels of detail, and we use drill-down to denote moving deeper in the aggregate hierarchy to reveal increasing detail [9], consistent with the "overview → zoom/filter → details-on-demand" mantra [22].

In interactive or computational contexts, hierarchical representations support tasks such as action segmentation and cross-modal video-text alignment, underscoring the value of multilevel structure [1, 26]. In instructional and exploratory settings, frameworks such as HETree and semantic windows operationalize hierarchy through chunking and level-wise navigation, improving exploration and retrieval [4, 7, 14]. However, to our knowledge, most prior systems assume user interactivity to access layered content, and it remains underexplored whether similar benefits can be achieved in non-interactive, fixed-pace videos where viewers cannot control pacing. Although many data-change operations exist, including aggregation/roll-up, filtering, and reconfiguration, we focus on drill-down transitions because they align with overview-to-detail traversal along a predefined hierarchy, which mirrors many real-world decision processes [9, 12, 26]. Recent evidence shows that adding narrative structure to visualization improves task efficiency and, for single-insight comprehension tasks, effectiveness compared with conventional visualizations, which strengthens the rationale for progressive drill-down disclosure in passive videos [21]. Finally, following chunking principles, we present information in manageable units, which is particularly pertinent to passive data-video formats [17].

3 Proposed Approach

3.1 Conceptual Model: Drill-Down Structuring

Building on prior work, we propose a conceptual model in which hierarchical drill-down structuring serves as segmentation and chunking for data videos, to guide the future design for optimising cognitive clarity.

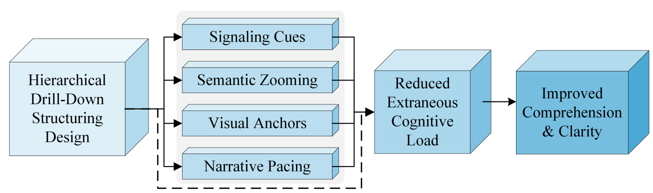

Figure 1: Expanded model of hierarchical drill-down structuring.

As shown in Figure 1, we posit this design enhances understanding primarily by reducing extraneous cognitive load. The model builds on a causal chain (design improvement → lower cognitive load → better comprehension) by adding four key design elements as mediators: signalling cues, semantic zooming, visual anchors, and narrative pacing (aligned with multimedia learning principles). Each element, based on multimedia learning theory and past research, helps direct viewer attention or organize information to reduce unnecessary mental effort. For example, signalling cues (visual highlights or pointers) focus attention on important parts, lessening the effort needed to search or filter content [16]. Semantic zooming gradually discloses data at increasing levels of detail, avoiding overload by matching the information density to the viewer's cognitive capacity [5, 9]. Visual anchors (persistent reference visuals or motifs) keep context as the story progresses, preventing disorientation and unnecessary processing [11]. Narrative pacing (the controlled speed of information flow) makes sure the video's rhythm matches the viewer's necessary processing speed [6]. These mechanisms help structure the information flow, and the model also suggests that there can be direct benefits on understanding from the well-designed structure itself. Overall, our model suggests that an effective drill-down structure can enhance understanding of complex content by reducing extraneous load and directing attention. Notably, fixed pacing rarely matches everyone's viewing habits, which poses practical challenges for experimental implementation.

3.2 Drill-Down Structuring Video Design

We created two short animated data videos on electric vehicle (EV) adoption and the state of New South Wales in Australia (NSW) rebate policy: a hierarchical drill-down version and a flat baseline. Both used the same data from the International Energy Agency and the Electric Vehicle Council. The drill-down video had three layers: a global overview, a focus on Australia comparing BEVs and PHEVs, and NSW local details. Both videos were silent to isolate structural effects. This ensured that participants' understanding and cognitive load were influenced only by the visual content and its structure, rather than by explanatory audio or emotional cues. This also reflects common scenarios where such videos may be played without sound (e.g., autoplay feeds, public displays) [24]. The baseline presented the same material in one continuous sequence without drill-down.

In the drill-down video, transitions between levels employed semantic zooming and animated object continuity: a visual element (e.g., an EV icon or the bar for Australia) persisted across scenes, changing size or position to signal the drill-down from global to national to state. Rather than abrupt cuts, smooth motion-based transitions guided the viewer from overview into details, preserving context and reinforcing relationships. For example, the national-level pie chart morphed into a localized bar chart, highlighting EV distribution in NSW and visually "zooming in" on the data—mimicking an interactive drill-down while the video remained non-interactive [11]. We also applied multiple visual encoding techniques to enhance interpretability: colour (consistent palette for categories, e.g., BEVs vs. PHEVs), position & layout (clear left-to-right, top-to-bottom arrangement), size & scale (proportional scaling to emphasize quantitative differences), and motion (smooth animations that show how one view transitions to the next).

3.3 Research Questions and Hypotheses

Guided by our model, we ask: RQ1—Does drill-down structuring improve factual recall relative to a flat baseline? RQ2—Does it reduce perceived cognitive load (NASA-TLX)? RQ3—Does it increase perceived comprehension of complex data? Accordingly, we posit: H1—Viewers of the drill-down video will recall more factual details than viewers of the flat baseline video. H2—Viewers of the drill-down video will report lower perceived cognitive load. H3—Viewers of the drill-down video will report higher perceived comprehension than baseline viewers.

4 Experiment

We conducted an online survey through Prolific with English-fluent adults living in Australia (N=100, 50 per condition) from various backgrounds. Participants provided informed consent. No personally identifiable information was collected, and no deception or invasive procedures were involved. Participants could withdraw at any time. Figure 2 presents the overview of the study procedure. (Study ID: 68574218797a3e6e4ea92fc5 / 6860b1eb2efb8555194d6683)

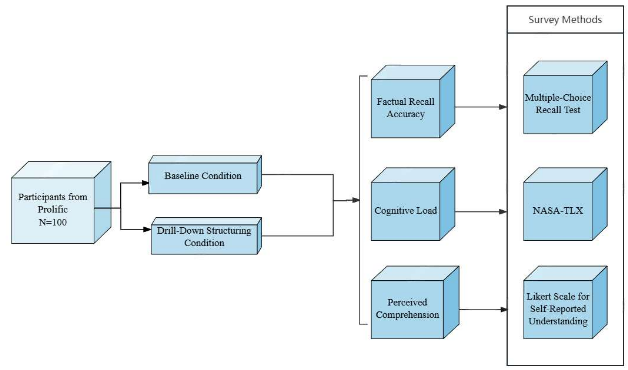

Figure 2: Overview of the study procedure.

4.1 Measurements

The post-video questionnaire collected objective and subjective measures: Factual Recall: Participants answered seven multiple-choice questions testing their recall of specific facts and figures presented in the video (e.g., what car models fully met NSW rebate criteria). A score (%) was computed based on correct answers. Perceived Cognitive Load: We adapted NASA-TLX to gauge mental effort and difficulty (Likert scale 1-7) [10]. Sample items included "The video's pace was overwhelming at times" and "I had to put a lot of mental effort into following the story." These targeted extraneous cognitive load. Perceived Comprehension: We used a 7-item scale on clarity and understanding (Likert 1-7). Sample items: "The video content was easy to understand" and "The structure of the video helped me retain key facts." Higher ratings indicate greater perceived clarity and understanding. Each study session lasted about 4-5 minutes.

4.2 Evaluation

Participants were assigned to view either a drill-down version or a baseline version of the same data video, and afterward, they completed a series of validated measurement tasks. We used two-tailed independent-samples t-tests (α = 0.05) and also reported the standardized effect size, Cohen's d, along with 95% confidence intervals (CI) for the difference in group means. The mean difference is defined as Δ = Drill-down — Baseline (recall in percentage points, pp; load and understanding in 1-7 raw units), and Cohen's d is signed similarly (positive values favour drill-down). We also summarized group variation using standard deviations (SD), expressed in the same units (pp for recall; 1-7 for load and understanding).

Recall Accuracy. As expected, factual recall was numerically higher in the drill-down condition compared to the baseline (42.3% vs. 40.7%; Δ = +1.6 pp), but this difference was not statistically significant, t (98) = 0.50, p = 0.617, 95% CI = [-4.8, +8.0] pp; Cohen's d = +0.10 (small, favouring drill-down). SDs were similar (15.5 vs. 15.2 pp), indicating comparable response variability. Therefore, while drill-down layering might help organize information, it did not produce a reliable improvement in factual retention under the current pacing and density.

Cognitive Load. Perceived load was numerically higher in the drill-down than in the baseline condition (M = 5.12 vs. 5.00; Δ = +0.12), t (98) = 0.67, p = 0.504, 95% CI = [-0.24, +0.48]; Cohen's d = +0.13 (small). SDs were comparable (1.18 vs. 1.13). Qualitative comments from both groups noted fast tempo and dense visuals, suggesting that these shared constraints likely elevated cognitive effort across conditions and diluted any benefit expected from segmentation.

Perceived Comprehension. Perceived comprehension was numerically lower in the drill-down than in the baseline condition (M = 3.22 vs. 3.47; Δ = -0.25), t (98) = -1.02, p = 0.310, 95% CI = [-0.74, +0.24]; Cohen's d = -0.20 (small, favouring baseline). The drill-down group also showed wider variance SDs (1.38 vs. 1.32), hinting at individual differences: some viewers found layering clarifying, while others experienced transitions and speed as disruptive.

5 Discussion and Limitations

The drill-down video showed marginally higher factual recall, but overall differences in outcomes were non-significant. Under fixed-paced, non-interactive conditions, drill-down structuring alone did not produce measurable cognitive benefits. These patterns suggest a complex trade-off: while drill-down design may offer structural benefits in theory, its effectiveness seems sensitive to implementation factors such as pacing, visual density, and user preferences.

A recurring theme from participant feedback was the fast pacing and high information density of both videos. Participants in the drill-down group noted: "The video was quality, but the amount of stats was a lot to take in," and "It probably could play a little slower to be better to digest and give more time to read through everything." Others commented, "Too fast. Too many facts," and "Too many figures in the videos. As someone who has a research background, the data was still overwhelming for me." In parallel, feedback from the baseline group echoed similar concerns: "The graphs were good to aid understanding, but it was way too fast." These comments reflect the inherent challenge of visual complexity and time-locked delivery: when dense quantitative information is presented rapidly without adaptive pacing or scaffolding, even a well-designed segment structure may not suffice to alleviate mental burden. Thus, one critical implication is that pacing flexibility may be a prerequisite for hierarchical drill-down to help viewers. Another implication is that narrative scaffolding (e.g., narration or explanatory text) and visual clarity may be needed alongside structuring. Some participants explicitly mentioned missing narration: "It needs narration and more interactive animations to help people understand the information." This suggests that combining drill-down with voice-over or text cues could reduce the interpretation effort and integrate segmented visuals into a coherent story. Besides, visual clutter should also be minimized: simplifying charts, using clear labels, and employing visual anchors can help guide attention and reduce load.

Notably, these findings also suggest new directions for future research. One promising direction is the integration of visual anchors—persistent or dynamic graphical elements that help orient the viewer across transitions and link segments into a cohesive visual schema. Specifically, in a public display context, drill-down should be paired with slower loop pacing, fewer concurrent visuals, and stronger persistent anchors to improve processing and comprehension. For personal viewing, AI-driven systems can adjust drill-down transitions in real time. By detecting likely overload from eye tracking or physiological signals, system can pause or replay segments at a slower pace. When basic controls are available, AI can personalize drill-down paths using adaptive narration and simple viewer controls. Although our study found no significant gains, qualitative feedback points to clear opportunities for future improvements.

Limitations: The sample size (N=100) was adequate to detect moderate effects, but the non-significant results suggest that any benefits of drill-down structuring may be subtle or context-dependent. Viewing was passive and fixed pace with no pause or rewind, which likely suppressed benefits and produced elevated load, as noted by participants. Removing narration and music shifted processing to the visual channel and may have increased extraneous load. Cognitive load and comprehension were self-reported and may miss real-time processing, so objective measures such as eye-tracking, physiological signals, and behavioural performance are needed. Both videos were visually dense with minimal scaffolding, which may have limited processing and retention.

6 Conclusion

This study examined whether drill-down structuring in linear, non-interactive, sound-off data videos improve understanding. We introduced a conceptual drill-down model for hierarchical storytelling and tested it in a controlled between-subjects experiment. Drill-down resulted in a small improvement in factual recall, but differences in recall, cognitive load, and perceived comprehension were not statistically significant. Participants' comments indicated that high information density and fast pacing often overshadowed structural benefits during processing. These findings suggest that structure alone is insufficient under consistent conditions. When deploying drill-down in linear, non-interactive, sound-off data videos, combine it with calibrated pacing, limited concurrent visuals, consistent visual anchors, and brief on-screen cues to enhance clarity. Future research should systematically evaluate anchor designs and adaptive pacing, including AI-assisted overload detection using gaze or physiological signals. In settings where minimal viewer controls are feasible, it should also test personalized drill-down paths to keep cognitive load manageable.

References

[1] Hyemin Ahn and Dongheui Lee. 2021. Refining action segmentation with hierarchical video representations. In *Proceedings of the IEEE/CVF international conference on computer vision*. 16302–16310. doi:10.1109/ICCV48922.2021.01599

[2] 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*. 1459–1468. doi:10.1145/2702123.2702431

[3] 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. doi:10.1145/3206505.3206552

[4] Jim Bastiras and Bruce H Thomas. 2017. Combining virtual reality and narrative visualisation to persuade. In *2017 International Symposium on Big Data Visual*

Analytics (BDVA). IEEE, 1–8. doi:10.1109/BDVA.2017.8114623

[5] Benjamin B Bederson and James D Hollan. 1994. Pad++: A zooming graphical interface for exploring alternate interface physics. In *Proceedings of the 7th annual ACM symposium on User interface software and technology*. 17–26. doi:10.1145/192426.192435

[6] Nicolas Biard, Salomé Cojean, and Eric Jamet. 2018. Effects of segmentation and pacing on procedural learning by video. *Computers in Human Behavior* 89 (2018), 411–417. doi:10.1016/j.chb.2017.12.002

[7] Nikos Bikakis, George Papastefanatos, Melina Skourla, and Timos Sellis. 2016. A hierarchical aggregation framework for efficient multilevel visual exploration and analysis. *Semantic Web* 8, 1 (2016), 139–179. doi:10.3233/SW-160226

[8] Ruochen Cao, Subrata Dey, Andrew Cunningham, James Walsh, Ross T Smith, Joanne E Zucco, and Bruce H Thomas. 2020. Examining the use of narrative constructs in data videos. *Visual Informatics* 4, 1 (2020), 8–22. doi:10.1016/j.visinf.2019.12.002

[9] Niklas Elmqvist and Jean-Daniel Fekete. 2009. Hierarchical aggregation for information visualization: Overview, techniques, and design guidelines. *IEEE transactions on visualization and computer graphics* 16, 3 (2009), 439–454. doi:10.1109/TVCG.2009.84

[10] 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. doi:10.1016/S0166-4115(08)62386-9

[11] Jeffrey Heer and George Robertson. 2007. Animated transitions in statistical data graphics. *IEEE transactions on visualization and computer graphics* 13, 6 (2007), 1240–1247. doi:10.1109/TVCG.2007.70539

[12] Jeffrey Heer and Ben Shneiderman. 2012. Interactive dynamics for visual analysis: A taxonomy of tools that support the fluent and flexible use of visualizations. *Queue* 10, 2 (2012), 30–55.

[13] Jessica Hullman, Steven Drucker, Nathalie Henry Riche, Bongshin Lee, Danyel Fisher, and Eytan Adar. 2013. A deeper understanding of sequence in narrative visualization. *IEEE Transactions on visualization and computer graphics* 19, 12 (2013), 2406–2415. doi:10.1109/TVCG.2013.119

[14] Alexander Kalinin, Ugur Cetintemel, and Stan Zdonik. 2014. Interactive data exploration using semantic windows. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*. 505–516. doi:10.1145/2588555.2593666

[15] Dongyang Liu. 2024. The effects of segmentation on cognitive load, vocabulary learning and retention, and reading comprehension in a multimedia learning environment. *BMC psychology* 12, 1 (2024), 4. doi:10.1186/s40359-023-01489-5

[16] Richard E Mayer and Roxana Moreno. 2003. Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist* 38, 1 (2003), 43–52. doi:10.1207/S15326985EP3801_6

[17] George A Miller. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review* 63, 2 (1956), 81. doi:10.1037/h0043158

[18] Chuan-Yu Mo, Chengliang Wang, Jian Dai, and Peiqi Jin. 2022. Video playback speed influence on learning effect from the perspective of personalized adaptive learning: A study based on cognitive load theory. *Frontiers in Psychology* 13 (2022), 839982. doi:10.3389/fpsyg.2022.839982

[19] Chris North and Ben Shneiderman. 2000. Snap-together visualization: can users construct and operate coordinated visualizations? *International Journal of Human-Computer Studies* 53, 5 (2000), 715–739. doi:10.1006/ijhc.2000.0418

[20] Edward Segel and Jeffrey Heer. 2010. Narrative visualization: Telling stories with data. *IEEE transactions on visualization and computer graphics* 16, 6 (2010), 1139–1148. doi:10.1109/TVCG.2010.179

[21] Hongbo Shao, Roberto Martinez-Maldonado, Vanessa Echeverria, Lixiang Yan, and Dragan Gasevic. 2024. Data storytelling in data visualisation: Does it enhance the efficiency and effectiveness of information retrieval and insights comprehension?. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–21. doi:10.1145/3613904.3643022

[22] Ben Shneiderman. 1996. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages*. 336–343. doi:10.1016/b978-155860915-0/50046-9

[23] Ben Shneiderman and Catherine Plaisant. 2006. Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. In *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*. 1–7. doi:10.1145/1168149.1168158

[24] Agnieszka Szarkowska, Valentina Ragni, Sonia Szkriba, Sharon Black, David Orrego-Carmona, and Jan-Louis Kruger. 2024. Watching subtitled videos with the sound off affects viewers' comprehension, cognitive load, immersion, enjoyment, and gaze patterns: A mixed-methods eye-tracking study. *PLoS One* 19, 10 (2024), e0306251. doi:10.1371/journal.pone.0306251

[25] Tamara Van Gog, Fred Paas, Jeroen JG Van Merriënboer, and Puk Witte. 2005. Uncovering the problem-solving process: cued retrospective reporting versus concurrent and retrospective reporting. *Journal of Experimental Psychology: Applied* 11, 4 (2005), 237. doi:10.1037/1076-898X.11.4.237

[26] Bowen Zhang, Hexiang Hu, and Fei Sha. 2018. Cross-modal and hierarchical modeling of video and text. In *Computer Vision – ECCV 2018*. 385–401. doi:10.1007/978-3-030-01261-8_23

[27] Zhenpeng Zhao and Niklas Elmqvist. 2023. The stories we tell about data: Surveying data-driven storytelling using visualization. *IEEE Computer Graphics and Applications* 43, 4 (2023), 97–110. doi:10.1109/MCG.2023.3269850

[28] Zhenpeng Zhao, Rachael Marr, Jason Shaffer, and Niklas Elmqvist. 2019. Understanding partitioning and sequence in data-driven storytelling. In *International Conference on Information*. Springer, 327–338. doi:10.1007/978-3-030-15742-5_32