Understanding Partitioning and Sequence in Data-Driven Storytelling

Zhenpeng Zhao,[1] Rachael Marr,[2], Jason Shaffer,[3] Niklas Elmqvist[1]

[1]University of Maryland, College Park; [2] Cisco Systems, Inc.; [3] U.S. Naval Academy
zhaoz@umd.edu, rachael.marr@gmail.com, tshaffer@usna.edu, elm@umd.edu

Abstract. The comic strip narrative style is an effective method for data-driven storytelling. However, surely it is not enough to just add some speech bubbles and clipart to your PowerPoint slideshow to turn it into a data comic? In this paper, we investigate aspects of *partitioning* and *sequence* as fundamental mechanisms for comic strip narration: chunking complex visuals into manageable pieces, and organizing them into a meaningful order, respectively. We do this by presenting results from a qualitative study designed to elicit differences in participant behavior when solving questions using a complex infographic compared to when the same visuals are organized into a data comic.

Keywords: Narrative visualization, comics, sequential art, data comics.
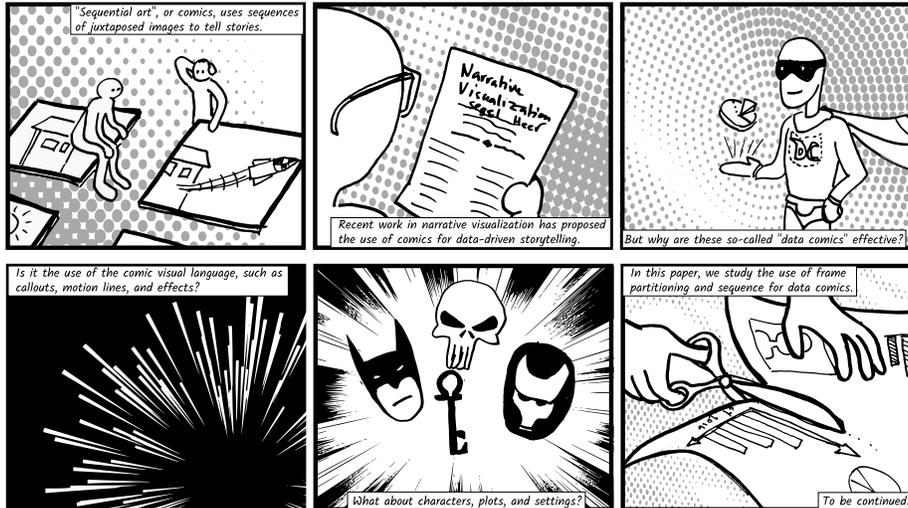
Fig. 1: Partitioning and sequence used as a method for organizing a complex visualization into the comic-strip narration style proposed by Segel and Heer [27].

1 Introduction

For many people, comics—stories told using a juxtaposition of illustrations, text, and visual annotations [6] that is also known as *sequential art* [24] or *sequential*

images [7]—draw their eyes and compel them to follow the sequence of frames on the page with an almost eerie power. These are the same people who, whenever they encounter the ubiquitous comic strip taped to an office door, feel the urge—often after a furtive glance around to ensure no one's watching—to stop and read the strip, chuckle softly to themselves, and then continue about their day. Part of the reason why comics have such compelling, even spellbinding, qualities is surely the combination of characters, visual effects, and drawing styles that provide a welcoming and familiar sense to the reader [5]. However, another reason is that comics organize and present plots, even complex ones, into mostly-linear sequences of frames that not only advance the story, but also guide and pull the reader into the narrative [24]. This mechanism is obviously not restricted to children's stories, superheroes, and funnies, so comics have also been harnessed for more "serious" applications, including statistics [15], physics [14], and even the study of comics itself [24]. In fact, Segel and Heer [27] listed "comic strip" as one of the seven genres of narrative visualization, and several recent papers have since advanced the notion of "data comics" to convey data using the visual language of sequential art [2, 3, 13, 33]. But why do data-driven comics work?

In this paper, we attempt to answer this question. While visual elements such as characters, visual language, and comic-style rendering no doubt contribute, and in fact may even strengthen the appeal and engagement of the data comic, their impact for a data-driven comic is likely secondary and cosmetic. Many data comics do not even use characters, visual effects, or freehand drawing to be effective [2, 3]. Based on the literature [24, 25], we instead focus on the use of the two primary storytelling mechanisms used in comic strip narration: (1) *partitioning*, or the organization of a complex story (i.e, dataset) into individual frames, and (2) *sequence*, or the ordering of frames into a narrative.

To achieve this, we conduct a qualitative study investigating the impact of partitioning and sequence by comparing comprehension and experience for a complex infographic and a comic representing the same data. Our results underline the importance of the partitioning and sequence mechanisms in making even complex and large-scale datasets understandable to a lay audience.

2   Related Work

Using comics for data-driven storytelling lies at the intersection of visual communication, storytelling, and sequential art. Below we review relevant prior art.

2.1   Visual Communication and Visualization

Visualization is a particular form of visual language traditionally used for solitary sensemaking. The notion of communication-minded visualization (CMV) [29] builds on ideas from visual communication by noting that visualization can often be used for more than just individual insights. Several examples of CMV systems exist, including Themail [30], the Baby Name Explorer [32], and Isis [26].

2.2 Visual Storytelling

Already in 2001, Gershon and Page [13] suggested using storytelling in visualization to improve its use for visual communication. Despite this, it is only recently that data-driven storytelling was fully embraced by the visualization community, with a survey by Segel and Heer in 2010 [27], and successful workshops at the annual conference in 2010 and 2011 [9, 10]. Hullman and Diakopoulos followed this up by studying how framing, context, and design impact the rhetoric of a narrative [16]. Since then, several practical methods and techniques have been proposed, including using sketching for narration [23], story points in Tableau [22], and automatic spatialization for visual exploration [19]. Most recently, Hullman et al. [17] studied sequence in narrative visualization, proposing a graph-driven approach for transitioning between views to minimize load on the viewer.

2.3 Comics for Visualization

Recent efforts have tried to harness comics for visualization. Jin and Szekely [20, 21] proposed a visual query environment that uses a comic-strip metaphor for presenting temporal patterns. However, their system solely uses comic strips for layout, and does not leverage the full potential of comics for communication.

The most recent and most relevant work in this vein is Graph Comics [2], which are data-driven comics used for telling stories about dynamic networks. More recently, Bach et al. [3] presented the generalized notion of data comics, but does not evaluate its partitioning and sequencing mechanisms. Knowledge can be maintained with proper temporal logic and intervals [1]. Interactive tools is able to validate the ranks and intervals of sequenced data [31].

3 Design Space: Data Comics

Data comics is a visual storytelling method based on sequential images consisting of data-driven visual representations. The motivation is to build engaging narratives about data through the familiar visual language of comics. Here we synthesize existing work on data comics, including in particular Bach et al. [3].

- **Basic Model:** We define a *comic* as a sequence of *panels* organized into one-dimensional *tiers* (or strips) and separated by *gutters*, or spacing, between the panels [12, 24]. The panels in a tier are organized to be read from left to right to form a narrative (at least in Western cultures). Tiers can in turn be organized into *pages*, where each tier becomes a row separated by a vertical gutter, and several pages can be linked together into a *book* (or comic book).
- **Panel Content:** Unlike a normal comic, most panels in a data comic consist of visualizations that convey information using graphical means.[1] These could be simple and familiar statistical graphics such as barcharts, time-series

---

[1] For engagement and effect, a few panels may consist solely of artistic content, but this puts corresponding artistic burden on the designer.

charts, and piecharts, or more advanced visualizations such as treemaps [28], node-link diagrams, or even parallel coordinate plots [18], all depending on the visualization literacy of the intended audience.

– **Characters, Annotation, and Effects:** A data comic would not be a comic if it did not also leverage comic elements:
   • **Comic-style rendering:** To emphasize the comic medium, content can be drawn using non-photorealistic rendering (e.g. [8]).
   • **Characters:** Characters may drive narrative even in a data-driven story.
   • **Comic elements:** Access to common elements used in comics, such as motion lines, highlights, or onomatopoeia (words that mimic sounds).
   • **Captions, speech, thoughts:** Visuals are often scaffolded by text in captions as well as speech and thought balloons [12, 24].

– **Layout Management:** The layout of a data comic—the organization of panels into tiers and pages—is an important consideration in creating a narrative. To facilitate easily constructing a narrative, the model should allow a designer to control layout.

– **Viewing:** Finally, after a data comic has been created, its purpose is to be viewed by its intended audience to convey its designer's story (and message). Just like a traditional comic, the default view for a data comic is to view an entire page, with all of the panels visible. Since screens are different from the written page, however, it also makes sense to support a single-panel navigation mode, where the viewer can sequentially navigate backwards and forwards in the comic. This is not unlike slideshows in PowerPoint.

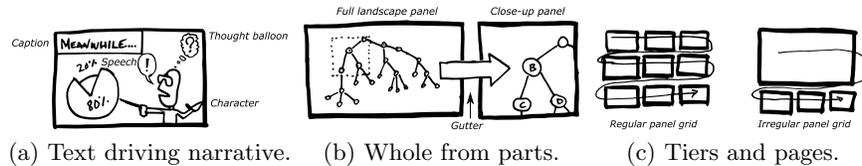(a) Text driving narrative.   (b) Whole from parts.   (c) Tiers and pages.

Fig. 2: (a) The use of text to drive narrative. (b) Encapsulation yield closure as the viewer connects them and sees the whole. (c) Panels organized into tiers.

4   Telling Stories Using Data Comics

Creating a data comic requires some knowledge of the narrative style of comics as well as the use of text in driving the story. Here we draw on the general literature of comics to operationalize these concepts for data comics [11, 24].

4.1   Comics Narration

While comics narration is generally linear, the medium by necessity cannot present the continuous narrative used in digital video or film. Instead, comics

narration depends on *encapsulation*, the focus on particular moments in the possible narrative arc to be represented in individual panels. The edges of the panels represent the limits of representation within this smaller bit of the page layout, but each panel also contributes to the larger unit, leading to an interpretive experience that is both linear and holistic since each panel is interpreted individually but also as part of the larger page. The gutters, while delineating the limits of each panel, also indicate the necessity for the reader to engage in *closure* in order to generate a more coherent narrative (Figure 2b).

While traditional comic creators typically work in a constructivist fashion, building scenes and sequences from the bottom up, a data comic may conceptually be designed in more of a top-down fashion: starting with the data and its manifestations. However, the mechanism of *partitioning* the subject matter into a hierarchy, the lowest tier of which is the panel, is the same across both applications of the comic medium. In other words, partitioning is concerned with the subdivision of story (or data) into comprehensive units, which continues until each unit is small enough that it can be conveyed in a single panel.

The creator then decides on a partitioning *sequence*—essentially, a hierarchy traversal—to generate the order of panels. The sequence of syntagmatically related panels can follow a number of patterns. It may be temporal, following the same event as it unfolds in time. It may be spatial or spatial/temporal, moving between a number of locations at a given moment or during a specified time, or else presenting a series of localized micro-events happening within a larger framework (akin to a series of reaction shots to the same event used in film). Panels could also relate conceptually, presenting a series of related abstractions and calling upon the reader to form associations between them.

4.2 Panel Content and Size

The comic panel is among the most fungible methods of representation, capable of presenting visual data ranging from a full landscape or urban horizon to a mid-range "street scene" shot to a close-up portrait shot or even the so-called extreme close-up of a particular detail. While layout possibilities are nearly limitless, most Western comics follow some variation of a grid pattern in which rows are read left to right while proceeding from the top to the bottom of the page.

The combination of larger and smaller panels can be used to convey a large quantity of visual information in a large panel while highlighting particular elements in smaller panels. For example, this could be done by presenting a landscape or a cutaway view of a house in a larger panel, and then presenting close-ups of visual elements in smaller panels collected near the larger panel.

4.3 Textual Narration

Comics panels may feature dialogue or textual narration, or these elements may be omitted, a stylistic choice embedded in the form's eighteenth-century origins. The satirical engravings of the British artist William Hogarth generally omitted dialogue, whereas politics cartoons dating from the eighteenth and nineteenth

centuries often featured multiple characters speaking (to all appearances simultaneously) in a one-panel image. Narration and dialogue are typically omitted for action sequences, focusing the reader's attention on the implied movement of the characters or objects in the panels. Dialogue is generally placed in speech balloons that are superimposed on the image. Narrative captions can be placed in a box and delivered in the first, second, or third person (Figure 2a).
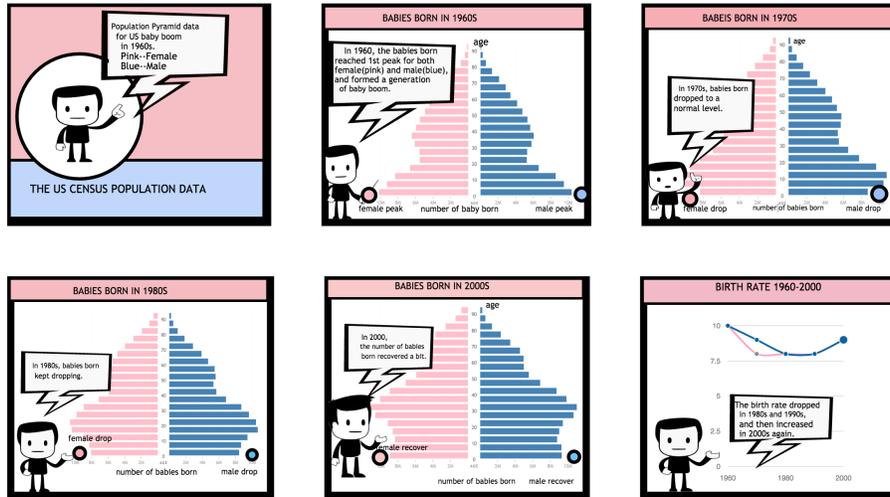

Fig. 3: Data comic of the U.S. baby boom of the 1960s created from public source material using our data comics authoring tool.

5   Implementing Data Comics

We have implemented a data comics authoring tool as a web application.[2] It is a hybrid application consisting of both client-side and server-side components. Client-side components are built using JQuery for DOM manipulation and D3 [4] for visualization. The content is stored on the server-side backend, implemented as a simple Python server communicating using JSON-RPC.

6   Study: Partitioning and Sequence in Storytelling

Our working hypothesis in this paper is that data comics provide a more effective way of telling stories than a single visualization. To empirically explore the

---

[2] Please see https://streamable.com/pw7xi for a video showcasing our tool. Note that this URL is anonymized, so does not break the confidentiality of the reviews.

virtues of this premise, we conducted a qualitative user study comparing multi-panel data comics with a single infographic-style visualization for the same data. Here we describe the methods and results from this evaluation.

(a) Original infographic.

(b) Infographic with panels and captions.

(c) Partition with panels, no captions.
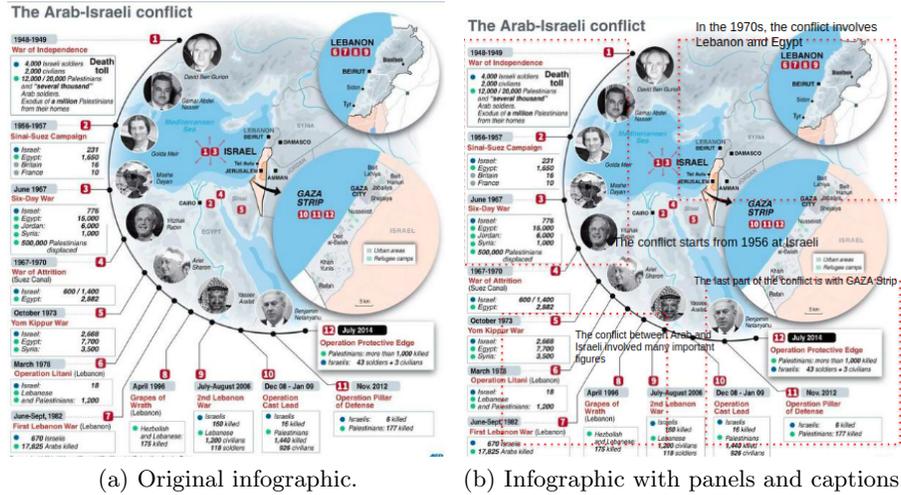
(d) Partition with panels and captions.

Fig. 4: (a) Original infographic. (b) Adding red dotted boxes to highlight panel locations. (c) The infographic is partitioned into panels. (d) The infographic is partitioned into panels and captions are added to structure the narrative.

– **Participants:** We recruited 12 paid participants (9 male, 3 female); all self-selected from the student population, aged between 20 and 31 years, with normal or corrected-to-normal vision, and proficient computer users.
– **Apparatus:** We conducted the experiment on a laptop computer equipped with a 15-inch 1280×800 LCD screen, a standard keyboard, and a three-button mouse. Both interface conditions were maximized to fill the screen.
– **Task and Datasets:** Each trial consisted of the participant using four types of data stories to answer questions: a single infographic versus a data comic, with and without captions for the different parts. Each story consisted of five or six panels to maintain the simplicity while providing sufficient information

for the story. Each story had an associated list of 7 to 9 questions designed to make the participant focus on details of the visualization. The stories included the Beer Origin Map (S_1), the Arab-Israeli Conflict (S_2), Smart Phishing Attacks (S_3), and World Wealthy People Distribution (S_4).

– **Metrics:** Our focus with the evaluation was to both study quantitative metrics, such as time and accuracy, as well as to collect subjective and qualitative feedback on the difference between data stories of data comics and original visualization. For this reason, we developed a questionnaire polling participants on their subjective experience of a story. This was administered to participants directly after each story, and consisted of 1–5 Likert-scale questions on engagement, speed, space efficiency, ease of use, and enjoyability.

– **Factors:** We included three factors in the experiment (all within-participants):
  - **Presentation (P)**: This factor modeled the presentation technique P given for solving questions: infographic (IG) or data comic (DC).
  - **Captions (C):** Access to partitions and captions. For the infographic, having access to the captions would show the bounding boxes of the partitions as well as the associated caption for each partition (Figure 4).
  - **Story (S)**: Modeling the impact of specific stories S on outcome.

– **Procedure:** An experimental session started with the participant arriving, reading and signing the consent form, and being assigned an identifier and story order. The administrator then explained the general goals and task. Each trial started with the administrator demonstrating how to read a data comic. The participant was then given four small examples (not the above stories), two with original visualizations (w/wt highlight of panel locations) and two with data comic (w/wt captions), and was allowed to ask questions about the examples and task during this time. When the participant finished training, they were given four stories, one at a time, opened in the appropriate tool and a paper sheet with questions. Participants saw each story once in sequence $(S_1, \ldots, S_4)$, with within-participant factors P and C chosen to counterbalance learning effects. They were given up to 10 minutes to answer the questions, and were encouraged to use all of the time. After answering, the participant was given the subjective questionnaire polling their experience in the trial. This was repeated for all four stories—one with the infographic, one with the infographic with captions, one with panels without captions, and one with panels with captions. A full session lasted approximately 50 minutes, including training and questionnaires.

6.1   Quantitative Results

Figure 5 depicts boxplots of the subjective ratings for the four types of tasks: a) infographic, b) infographic with highlights of panel focus locations and captions, c) data comic without captions, and d) data comic with captions. The ratings are for following effects on a 5-point Likert scale: engagement, speed, space-efficiency, ease of use, and enjoyability (Q1 through Q5). We analyzed the 5-point Likert scale of subjective ratings for effects of the technique P (infographic vs.
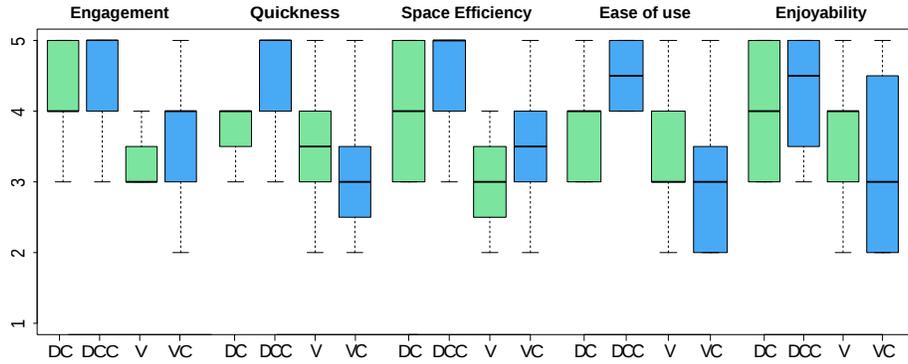
Fig. 5: Comparison between single infographic (V), infographic with caption (VC), data comic panels without captions (DC) and data comic panels with captions (DCC) of subjective ratings (Likert 1-5 scale).

data comic) and captions C (no captions vs. with captions), and found that the engagement (Q1), speed (Q2) efficiency (Q3), and enjoyability (Q4) were significantly different between the four techniques (Friedman tests, $p < .05$), but enjoyability (Q5) had no significant difference (Friedman tests, $p = .12$). We also found no significant effect of story S on any of the metrics.

6.2   Qualitative Feedback

Inviting Reading: Nine participants noted that the comic layout helped them view the materials as a whole story without explicit instruction.

Easy to Follow: Reading a complex infographic can be challenging. Partitioning the visualization into panels can help the reader follow the sequence of panels and captions to generate a thread. One participant mentioned that the Arab-Israeli conflict infographic is overwhelming at first glance, but that the first data comic panel is a great summary with supplementary information partitioned to other panels. Another participant noted that the data comic allows for skipping panels when reading while still following the big picture of the story.

Facilitating Focus: The data comic panels are organized in a sequence following a narrative. The audiences' attention is directed by the panel, so that the important information is contained in certain panels. One participant mentioned that during the user study, he was able to easily locate a couple of panels whenever he needed a certain kind of information. Another participant mentioned that having captions and panels is like having labels for the whole visualization.

Facilitating Memory: In our study, several participants mentioned that the panels suggested a structured way of reading through the sequence of the panels. Five participants noted that reading the data comic panels helped them remember the information when answering questions.

7   Discussion

Our qualitative evaluation indicated that data comics, especially with captions were significantly more engaging, space-efficient, faster and easier to use than original visualization/infograhic. The feedbacks from the participants also indicate that panels of partitions help focus and memory. The captions are particularly helpful when following the story and remembering the details. The participants mostly felt that the data comic was more effective that it invites reading and helps build up the story. The sequence of the panels in data comics are important especially when helping the participants recall detailed information on one of the panels with the help of captions. The overall sequence is more important than sequence of a small range, i.e. two panels about the topics parallel to each other can be changed without harming the whole storyline.

We explicitly chose not to measure time or correctness. There is likely little difference between data comics versus single visualization/infographic, and this perception was also confirmed by participants in our experiment. Rather, the strength of data comics comes from its approachable, compelling, and intuitive format. This is further validated by Lee et al. [23], who only collected subjective ratings from participants in their SketchStory evaluation.

Our work in this paper has several limitations. First of all, much of our argumentation of using sequential art for data is based on two assumptions: that the audience has (a) prior experience, and (b) a favorable opinion about comics. With no prior experience, much of the benefit of an established common ground in the visual language of comics is lost. Furthermore, given the sometimes questionable respectability of comics [12, 24], its use as a communication medium may be problematic. For example, it can be argued that a data comic may not be the best vehicle for presentations in very formal settings, such as a boardroom meeting. Similarly, the intrinsically light-hearted nature of comics may be inappropriate for sensitive or difficult topics, such as natural disasters, emergency situations, and other types of crises or stories on the loss of lives or livelihoods.

8   Conclusion and Future Work

We have given a background on storytelling using data comics that is grounded in research on sequential art. We have built a data comic authoring system to investigate this topic. We presented a qualitative study designed to elicit differences in participant behavior when solving questions using a complex infographic compared to when the same visuals are organized into a data comic. The results show that the sequences of partitions help focusing and memory.

Our future work will be continuing exploring narrative visualization using sequential art. For example, there may be situations when it is appropriate to combine dynamic visualizations with a static comic. Furthermore, we also want to incorporate more collaborative features into the system.

References

1. Allen, J.F.: Maintaining knowledge about temporal intervals. Commun. ACM **26**(11), 832–843 (Nov 1983)
2. Bach, B., Kerracher, N., Hall, K.W., Carpendale, S., Kennedy, J., Riche, N.H.: Telling stories about dynamic networks with Graph Comics. In: Proceedings of the ACM Conference on Human Factors in Computing Systems. pp. 3670–3682 (2016). https://doi.org/10.1145/2858036.2858387
3. Bach, B., Riche, N.H., Carpendale, S., Pfister, H.: The emerging genre of data comics. IEEE Computer Graphics and Applications **37**(3), 6–13 (2017). https://doi.org/10.1109/MCG.2017.33
4. Bostock, M., Ogievetsky, V., Heer, J.: D3: Data-driven documents. IEEE Transactions on Visualization & Computer Graphics **17**(12), 2301–2309 (2011)
5. Christiansen, H.C.: Comics and film: a narrative perspective. In: Magnussen, A., Christiansen, H.C. (eds.) Comics & culture: Analytical and Theoretical Approaches to Comics, pp. 107–122. Museum Tusculanum Press, University of Copenhagen (2000)
6. Cohn, N.: Beyond speech balloons and thought bubbles: The integration of text and image. Semiotica **197**, 35–63 (2013). https://doi.org/10.1515/sem-2013-0079
7. Cohn, N.: The Visual Language of Comics: Introduction to the Structure and Cognition of Sequential Images. Bloomsbury, London (2014)
8. Collomosse, J.P., Rowntree, D., Hall, P.M.: Rendering cartoon-style motion cues in post-production video. Graphical Models **67**(6), 549–564 (2005). https://doi.org//10.1016/j.gmod.2004.12.002
9. Diakopoulos, N., DiMicco, J., Hullman, J., Karahalios, K., Perer, A.: Telling stories with data: The next chapter—a VisWeek 2011 workshop (2011)
10. DiMicco, J., McKeon, M., Karahalios, K.: Telling stories with data—a VisWeek 2010 workshop (2010)
11. Duncan, R., Smith, M.J.: The Power of Comics: History, Form, and Culture. Continuum, New York, NY (2009)
12. Eisner, W.: Graphic Storytelling and Visual Narrative. W. W. Norton & Company (2008)
13. Gershon, N.D., Page, W.: What storytelling can do for information visualization. Communications of the ACM **44**(8), 31–37 (2001). https://doi.org/10.1145/381641.381653
14. Gonick, L., Huffman, A.: The Cartoon Guide to Physics. HarperPerennial, New York (1990)
15. Gonick, L., Smith, W.: The Cartoon Guide to Statistics. HarperCollins, New York (1993)
16. Hullman, J., Diakopoulos, N.: Visualization rhetoric: Framing effects in narrative visualization. IEEE Transactions on Visualization and Computer Graphics **17**(12), 2231–2240 (2011). https://doi.org/10.1109/TVCG.2011.255
17. Hullman, J., Drucker, S.M., Riche, N.H., Lee, B., Fisher, D., Adar, E.: A deeper understanding of sequence in narrative visualization. IEEE Transactions on Visualization and Computer Graphics **19**(12), 2406–2415 (2013). https://doi.org/10.1109/TVCG.2013.119
18. Inselberg, A.: The plane with parallel coordinates. The Visual Computer **1**(2), 69–91 (1985). https://doi.org/10.1007/BF01898350
19. Javed, W., Elmqvist, N.: ExPlates: Spatializing interactive analysis to scaffold visual exploration. Computer Graphics Forum **32**(3pt4), 441–450 (2013). https://doi.org/10.1111/cgf.12131

20. Jin, J., Szekely, P.A.: QueryMarvel: A visual query language for temporal patterns using comic strips. In: Proceedings of the IEEE Conference on Visual Languages and Human-Centered Computing. pp. 207–214 (2009). https://doi.org/0.1109/VLHCC.2009.5295262

21. Jin, J., Szekely, P.A.: Interactive querying of temporal data using a comic strip metaphor. In: Proceedings of the IEEE Symposium on Visual Analytics Science and Technology. pp. 163–170 (2010). https://doi.org/10.1109/VAST.2010.5652890

22. Kosara, R.: Story points in Tableau Software. Keynote at Tableau Customer Conference (Sep 2013)

23. Lee, B., Kazi, R.H., Smith, G.: SketchStory: Telling more engaging stories with data through freeform sketching. IEEE Transactions on Visualization and Computer Graphics **19**(12), 2416–2425 (2013). https://doi.org/10.1109/TVCG.2013.191

24. McCloud, S.: Understanding Comics: The Invisible Art. William Morrow Paperbacks (1994)

25. McCloud, S.: Comics: A medium in transition. Computer Graphics Forum **30**(3), xiii (2011)

26. Phan, D., Paepcke, A., Winograd, T.: Progressive multiples for communication-minded visualization. In: Proceedings of Graphics Interface. pp. 225–232 (2007). https://doi.org/10.1145/1268517.1268554

27. Segel, E., Heer, J.: Narrative visualization: Telling stories with data. IEEE Transactions on Visualization & Computer Graphics **16**(6), 1139–1148 (2010). https://doi.org/10.1109/TVCG.2010.179

28. Shneiderman, B.: Tree visualization with tree-maps: A 2-D space-filling approach. ACM Transactions on Graphics **11**(1), 92–99 (Jan 1992). https://doi.org/10.1145/102377.115768

29. Viégas, F., Wattenberg, M.: Communication-minded visualization: A call to action. IBM Systems Journal **45**(4), 801–812 (2006). https://doi.org/10.1147/sj.454.0801

30. Viégas, F.B., Golder, S., Donath, J.: Visualizing email content: portraying relationships from conversational histories. In: Proceedings of the ACM Conference on Human Factors in Computing Systems. pp. 979–988 (2006). https://doi.org/10.1145/1124772.1124919

31. Wang, T.D., Plaisant, C., Quinn, A.J., Stanchak, R., Murphy, S., Shneiderman, B.: Aligning temporal data by sentinel events: Discovering patterns in electronic health records. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 457–466. ACM, New York, NY, USA (2008)

32. Wattenberg, M.: Baby names, visualization, and social data analysis. In: Proceedings of the IEEE Symposium on Information Visualization. pp. 1–7 (2005). https://doi.org/10.1109/INFVIS.2005.1532122

33. Zhao, Z., Marr, R., Elmqvist, N.: Data comics: Sequential art for data-driven storytelling. Tech. Rep. HCIL-2015-15, Human-Computer Interaction Laboratory (2015)