Creating Data Comics for Data-Driven Storytelling

Zezhong Wang

Doctor of Philosophy

Institute for Language, Cognition and Computation
School of Informatics
University of Edinburgh

2022

Abstract

This doctoral thesis investigates research in understanding and creating data comics for data-driven storytelling. Data comic is a novel genre aiming to communicate insights from data through visualizations. Inspired by how audiences read and understand comics, data comics integrate texts and images in the 2D display, map time into space through panels and their ordering. Data comics provide the freedom of space-oriented layout of infographics and annotated charts while supporting the linear narration of videos and live presentations. They open up a potential for expressive storytelling with data visualizations. However, as a new genre of data-driven storytelling, the effectiveness, merits, and drawbacks of data comics are still unknown. Creating data comics is challenging, which requires multiple skills ranging from understanding data, presenting data with visualization techniques, storytelling techniques, drawing, literature to interaction and user-centered design approaches.

This thesis investigates two primary questions—*Is data comics an effective genre for data-driven storytelling? How to create effective data comics?* To answers the two questions, the author creates data comics, distills design patterns, refines creation methods, conducts controlled studies, workshops, collaborative design sessions, and interviews with experts from data visualization, statistics, professional illustrators, industrial visualization practitioners, students from design and data science.

Toward understanding and creating data comics for effective communication, this thesis presents (i) a Data Comic Creation Model featuring a structured working process of creating data comics, and as a theoretical framework to summarize the contribution of this thesis; (ii) a lab and an in-the-wild study comparing the effectiveness and engagement of data comics, infographics, and illustrated text; (iii) a structured data comic creation workshop with reusable materials and activities; (iv) data visualization cheat sheets supporting for using and learning visualization techniques; (v) applying data comics in reporting controlled user studies in Human-Computer Interaction; (vi) using interactive data comics for presenting non-linear data stories. At last, the thesis is finished by summarizing each chapter and planned future works.

This thesis highlights that data comics as one of the crown jewels of data-driven storytelling are promising for data-driven storytelling. We hope the knowledge, approaches, and techniques presented in this thesis can inform the practice and research of data comics in the future.

Acknowledgements

This project would not have been possible without the supports and inspiration from my supervisors, collaborators, examiners, reviewers, families, and friends during my PhD.

First, I'd like to express my deepest thanks to my principal supervisor Benjamin Bach, for introducing me to the world of data visualization and data-driven storytelling, guiding me through each stage of the process, not only teaching me the expertise of doing research but also encouraging me to discover my potential that boosted my confidence. I am also inspired by his passion for research and generous personality, I am grateful for his time, patience, and insight that helped me on track.

I would like to extend my deepest gratitude to my assistant supervisor Dave Murray-Rust, who has been my mentor since my master's degree and inspired me by his invaluable knowledge of doing HCI research and academic writing. I thank him for being supportive from beginning to the end during my study at the University of Edinburgh and helping me with each project.

I am very grateful to John Lee and Niklas Elmqvist for being my examiners, committing time, thoughts, and giving feedback to this thesis; Maria Wolters, for being the external reviewer of my PhD annual report, especially for nudging me to find the road map of my PhD; panelists (Melanie Tory, Miriah Meyer, Uta Hinrichs, Wolfgang Aigner) and candidates in VIS 2019 Doctoral Colloquium for giving constructive suggestions for my PhD.

I would also like to express my gratitude to my collaborators who made this thesis together with their high degree of expertise: (alphabetical order) Fanny Chevalier, Hugo Romat, Jacob Ritchie, Jingtao Zhou, Lovis Sindin, Matteo Farinella, Nathalie Henry Riche, and Shunming Wang.

I must also thank Sheelagh Carpendale for offering the internship opportunity and introducing me to join an exciting project at the end of my PhD, which provided me a novel view of the potential of sketches. I am looking forward to continuing our collaborations.

I'd also like to acknowledge my PhD scholarship offered by *University of Edinburgh*, *School of Informatics*, staff and nice colleagues from *Vis Hub* and *Design informatics* for being supportive to my research and bringing diverse perspectives.

Finally, I would like to thank my parents and grandparents for their unconditional love, and my friends for their accompanies that make my PhD not a lonely journey.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Zezhong Wang*)

Table of Contents

Chapter 1

Introduction

1.1 Background and Motivation

Working with data is crucial in the modern world. Data contains information that helps organizations and individuals to identify new opportunities and make informed decisions. Unfortunately, discovering evidence-based insights from data is an expensive process including a series of tasks such as data collecting, cleaning, visualizing, and analyzing [Philip Chen and Zhang, 2014]. To improve the accessibility and proficiency for working with data, an increasing number of tools have emerged, e.g., Tableau [1] and Microsoft Power BI [2], supporting users to transfer data into insights through simple actions. However, the process of data analysis is meaningless if the discovered insights fail to communicate to collaborators, stakeholders, the public, or other intended receivers [Cook and Thomas, 2005]. As one of the communicating approaches, *data visualizations* including statistical graphics, charts, infographics are increasingly used in the wild to deliver messages supported by data [Tufte, 1985].

Communicating with data visualizations takes thought and effort. Just showing the visualizations used in the exploration stage for a communication purpose may render the important information ignored and the most compelling story completely underwhelming. Especially when the visualization techniques are not familiar for the audience (e.g., Whiskers Plots (Box-Plots) and Time Curves [Bach et al., 2016b]), which require some degree of learning and understanding from the audience about the visual mappings, graphical conventions, symbols, marks, layouts, and patterns to properly decode the representation [Börner et al., 2016, Lee et al., 2016, Maltese et al., 2015]. To increase the accessibility of the insights discovered by data analysts, *effective communication* is desired.

[1] https://www.tableau.com/
[2] https://powerbi.microsoft.com/

1.1.1 Data Visualizations Used for Communication

In general, there are two categories of data visualizations: exploratory data visualizations used for data analysis, and explanatory data visualizations used for communication [Iliinsky and Steele, 2011]. Exploratory data visualization techniques facilitate data analysts to discover patterns, trends, clusters from data, while explanatory data visualizations are designed to deliver the messages and insights discovered from the data analysis phase to the audience. The potential audience of explanatory data visualizations can be broad, from experts to novices, e.g., high school students, industrial stakeholders, a grant committee, or the general public. In the communication process, data visualization techniques are intended to make information comprehensible for the audience through visual representation. Therefore, to create visualizations for effective communication, creators need to consider not only audiences' pre-knowledge of a certain domain, but also data visualization literacy – the ability to "interpret patterns, trends, and correlations in visual representations of data" [Börner et al., 2016]. Moreover, instead of analytical driven visualizations, visualizations for more casual audiences are made with effort and design consideration to improve engagement and memorability [Pousman et al., 2007]. E.g., adding embellishments [Bateman et al., 2010, Borkin et al., 2013] and presenting data with stories.

1.1.2 Data-Driven Storytelling

Rather than just "visualizing" data, data-driven storytelling involves interpreting the insights with data visualizations, connecting to the audiences' background, interests and presenting the content in a form appropriate to the message and the attention span of the audience [Riche et al., 2018]. Data-driven stories apply in a variety of settings for diverse audiences. Kosara and Mackinlay [2013] describe three typical scenarios: "self-running presentations for a large audience", e.g., news agencies like the Guardian, New York Times, Washington Post, and Bloomberg, investigating data-driven news to communicate and disseminate information and insights from data to the public [Gray et al., 2012]; "live presentation", e.g., Hans Rosling's Gapminder presentation in the 2006 TED conference [3]; "individual or small-group presentations" particularly indicate the interaction between the presenter and the audience, e.g., asking and answering questions that are not part of the presented story.

To adapt to the different communication scenarios and types of audience, the art of data-driven storytelling is not only about the content of a story, but also the *genres* and *techniques* of presenting the stories. Segel and Heer [2010] suggests a set of *genres* for presenting visual data stories, each genre has its distinct feature regarding layout, reading sequence, and frames

[3] https://www.ted.com/talks/hans_rosling_the_best_stats_you_ve_ever_seen#t-233313

presented at a time. Moreover, there are various *narrative techniques* such as element highlighting, linking through interaction, and tooltips that enable information navigation, explanation on-demand, and self-exploration [Stolper et al., 2016]. There is no one best genre or a storytelling technique working for all cases, the decision depends on scenarios, intended audience groups, and mediums.

1.1.3 Data Comics

Inspired by the way people read and understand comics [McCloud, 1993, Groensteen, 2007, Caldwell, 2012], *data comics* have come into focus as a highly flexible format that combines aspects of infographics and posters (i.e., spatial layout, elements with different sizes) with narrative aspects of videos (i.e., narration, storytelling pace, linear perceiving order, and surprise) [Zhao et al., 2015b, Bach et al., 2017a, 2016a].

The potential of using comics for data-driven storytelling lies in the unique features of comics. In comics, the integration of images and texts provides complementary content to achieve an effective and engaging presentation. To present the textual and visual contents, comic panels can break the complexity of information and explanations into a sequence of panels. This breakdown enables controlling the amount of information to be perceived at a time, and audiences can read at their speed. Panels that place in sequence lead to an explicit reading order to make sure the audience is always guided through the story. Not only that, the linearity provides an inherent condition for presenting temporal data. The panels placed in linear order work like keyframes in a video, while the layout of panels and contents provide a reach dimension that videos do not have. With carefully designed panel layouts, the creator could group contents, for instance, using rows and pages to structure a story. Commonly found in comics, metonymy employs images, words, or concepts from a different domain to foster the understandability of a concept from a target domain [Davies, 2019, Kovecses, 2010]. The visual languages in comics provide a rich collection for visual explanation [Cohn, 2013b], metaphor and embellishment are commonly used in comics, which is potentially supportive for understanding and recalling the messages with data visualizations [Borkin et al., 2013, Bateman et al., 2010]. Moreover, comics are highly adaptable and accessible as they do not require any advanced techniques to display, they could be present through various mediums such as posters, textbooks, small booklets, screens of different sizes. Comics can be created in many ways, e.g., by hand drawing, through design tools, or editing from computer-generated graphs.

For a long time, comics have been applied to communicate scientific phenomenon [Farinella, 2018] (e.g., Figure 1.1), demonstrate instructions (e.g., Figure 1.2), and promote public en-

gagement [Negrete and Lartigue, 2004]. Empirical evidence also supports the use of comics in teaching statistics [Takahashi et al., 2008, Gonick and Smith, 1993, Peck et al., 2011] and health-related communication, and found to be more engaging and easier to comprehend, especially for non-expert readers [Delp and Jones, 1996, Diamond et al., 2016].

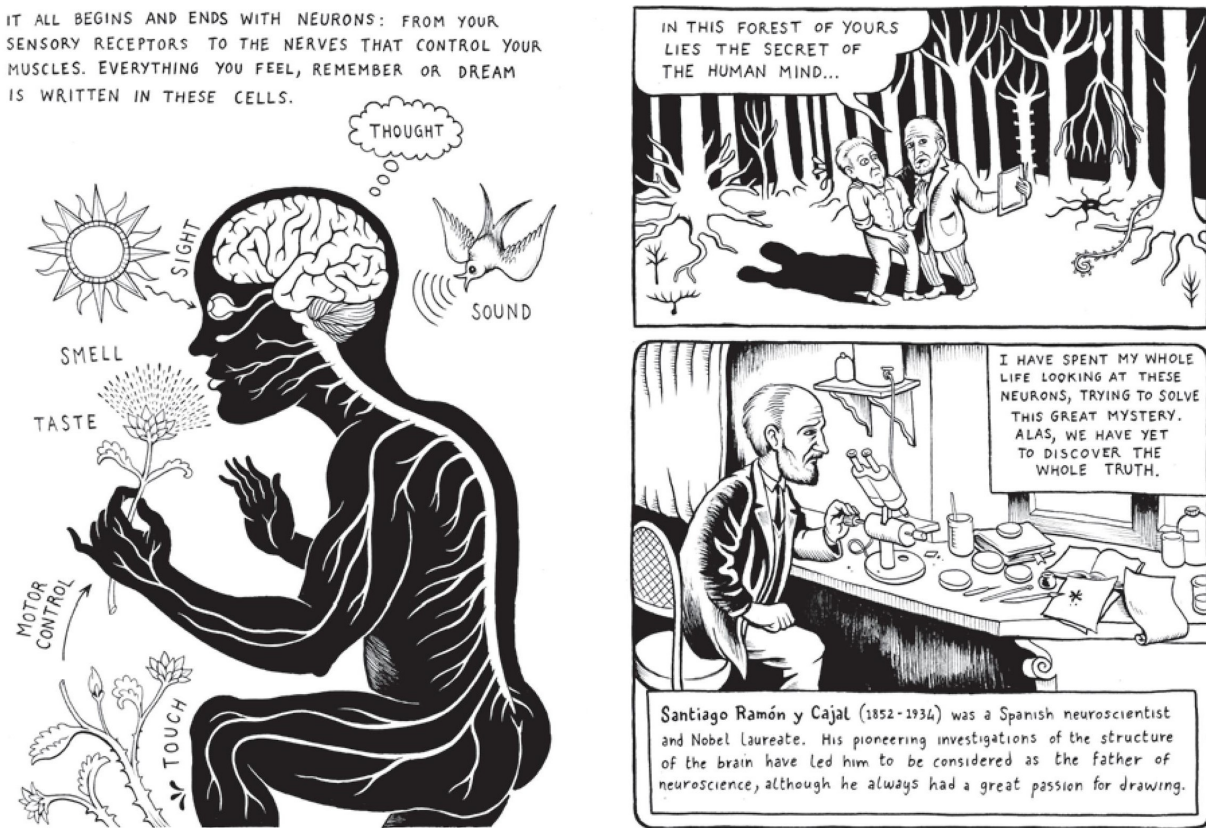


Figure 1.1: Neurocomic [Roš and Farinella, 2013] presents a journey through the human brain by using metaphor such as a place of neuron forests, memory caves, and castles of deception.

Before this PhD project, few works were found that use comics integrated textual and visual contents to explain data-based facts. For example, in Figure 1.3, French civil engineer Minard used three panels to illustrate the impact of the American Civil War on the European cotton trade, which helps traffic engineers to predict demand on existing or projected routes [Rendgen, 2018]. Nature journal published a 9-page example on the challenges of climate change [Monastersky and Sousanis, 2015]. The comic (Figure 1.4) presents an illustrative and informative story, it uses hand-drawing style sketches and statistical charts to demonstrate a series of key concepts, such as greenhouse effect, data-based facts about global temperature trend, predicted sea level, emission target by countries, and the changes of CO2 concentration over time. Other examples include the graph comic by Bach et al. [2016a] and eight examples the author of this thesis created as part of Master's degree dissertation [Bach et al., 2018c]. This thesis covers the

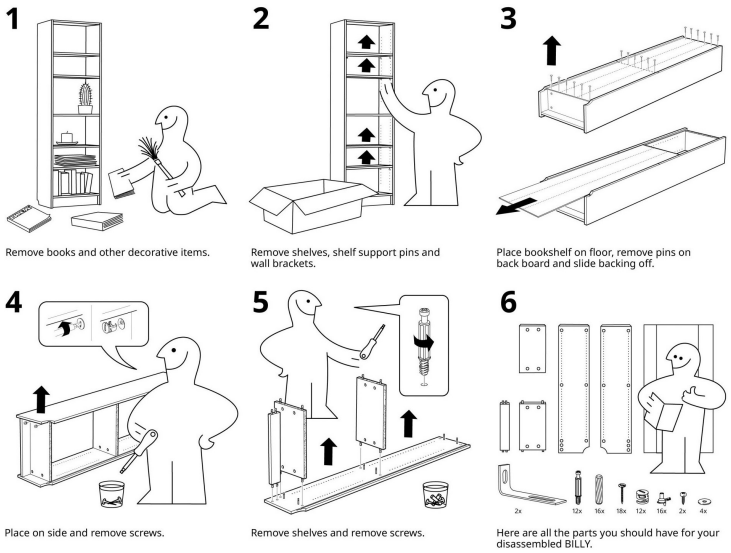


Figure 1.2: An example of IKEA's disassembly instructions presented in comic fashion.

emergence of the form and connects to the widest range of data comics to date (collected at `https://datacomics.github.io/`).

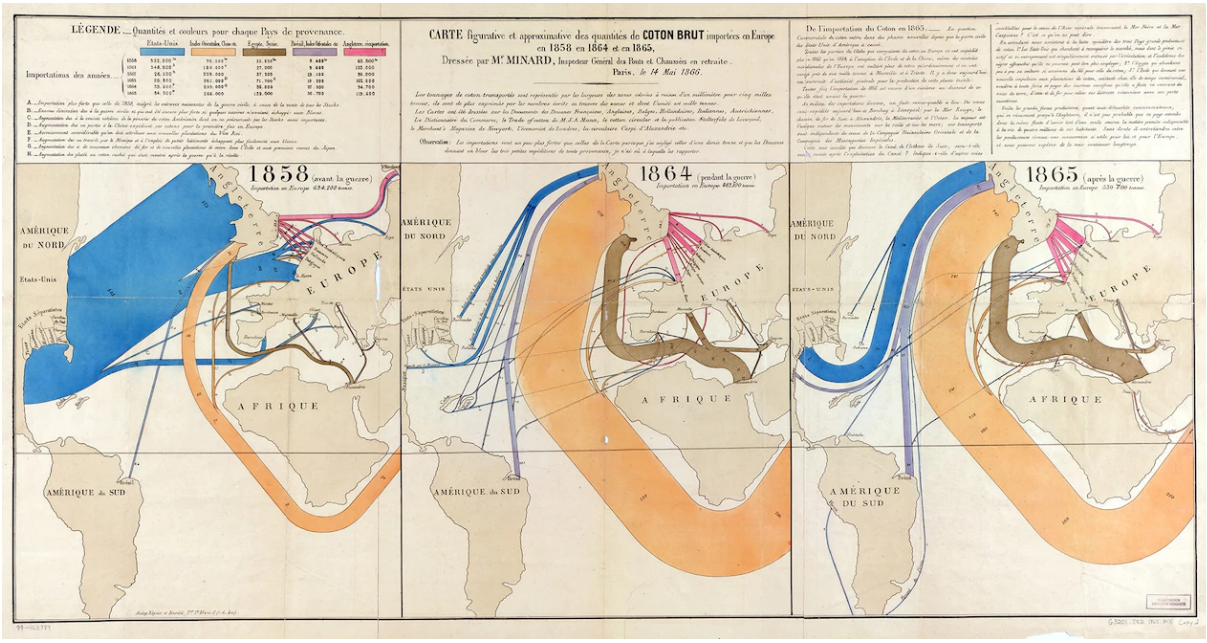


Figure 1.3: The maps created by Charles-Joseph Minard in 1866, it contains three panels to depict the temporal changes of origin and amount of cotton imported into Europe in 1858, 1864, and 1865.

Creating data comics can be a complex process. Gershon and Page [2001] noted that crafting data stories requires a diverse set of skills, which includes technical knowledge of computing engineering and data science, story directing, designing visual media, and a mind that "caters

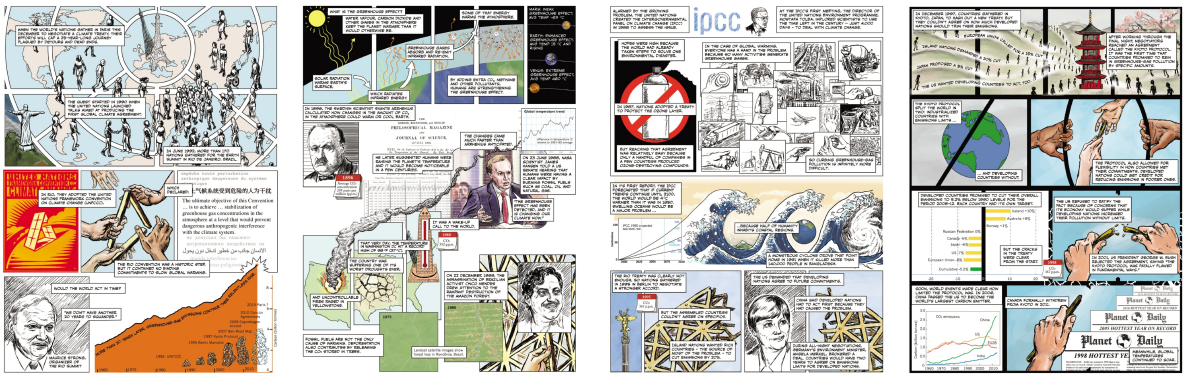


Figure 1.4: The first 4 of 9 pages of a comic examines the 25-year quest for a climate treaty that published in Nature.

to other modes of human information processing and thinking". Creating data comics requires creators to handle the transition, layout, and visual languages of comics to serve their communication objectives with data. The existing supports and knowledge for creating data comics is sparse, including data comic design patterns for different narrative purposes [Bach et al., 2018c], and DATACOMICSJS by Zhao et al. [2015b] for creators to embellish data visualizations with comic elements. However, there is no evidence about the effectiveness of using data comics, creating data comics with various types of data that work for diverse scenarios still requires a lot of manual work, and generally entails laboriously switching between data visualization tools and illustration tools [Bigelow et al., 2017]. At an earlier stage of exploring data comics, this thesis investigates two primary questions:

- ***Is data comic an effective genre for data-driven storytelling?***
- ***How to create effective data comics?***

1.2 Research Questions and Approaches

The two primary questions set motivation throughout this PhD, to make the questions more concrete and actionable, this section introduces the seven research questions (RQ1 to RQ7) and approaches presented in this thesis.

RQ1) Are data comics more effective compared to other data-driven storytelling formats?

Generally, the use of comics in classroom education and health-related communication appears to improve the reader's comprehension and engagement. However, how effective comics

are used for presenting data stories is still not known. As a step towards better understanding data comics, we provided the first structured investigation into the effectiveness of data comics.

To answer RQ1, this thesis starts with comparing data comics and the two closest popular alternatives: infographics and texts accompanied by visualizations (illustrated texts). Chapter 4 describes a comparison between these techniques in a controlled lab study with 36 participants reading the same three stories in a different format. Empirical and subjective data were collected on understandability, recall, preferences, and engagement. To complement this controlled study, we observed 50 groups of visitors engaging with comics and infographics in an open public space.

RQ2) What are the merits and drawbacks of data comics?

There are numerous differences between illustrated texts (i.e., magazine style), infographics, and data comics. These three static genres varied by the degree to which an explicit reading order is given, and how close the integration of text and picture is. Under the frame of these two dimensions, what are the merits and drawbacks of high linearity and close text-picture integration?

The results from Chapter 4 suggest that data comics improve understanding for most cases, and were generally rated high for enjoyment and engagement. To further understand what makes data comics preferred in general and disliked in some cases, we interviewed participants in the lab and in the wild.

RQ3) How to streamline the process of creating data comics?

The results in chapter 4 encourage the use of data comics for data-driven storytelling, from RQ2 to RQ3, we transfer our focus from understanding to creating data comics. Creating data comics requires multiple skills and through many phases. To provide streamlined approaches and tool supports, we looked into the workflow and challenges of the creating process.

Informed by the process of transforming data into visually shared stories, a working model described by Lee et al. [2015] (Figure 1.5), and data comic workshops conducted before, we redesign and run a data comic creation workshop (Chapter 5). The workshop was part of a 3-week assignment in an undergraduate illustration class, which includes a fast-forward session, creating data comics with self-prepared data, and group tutorials. We also identified challenges during the participants' creation process.

RQ4) How to effectively explain data visualization techniques in data comics?

Data visualization take an essential part in data comics. Effective communication with data comics requires the visualizations to be understood by the target audience. The workshop in chapter 5 suggests participants struggling with using visualization techniques, both for visualizing data and explaining the visualizations to the audience in the comics. Interpreting data visualizations is a general challenge for both data comic authors and audiences, especially with complex visualization techniques.

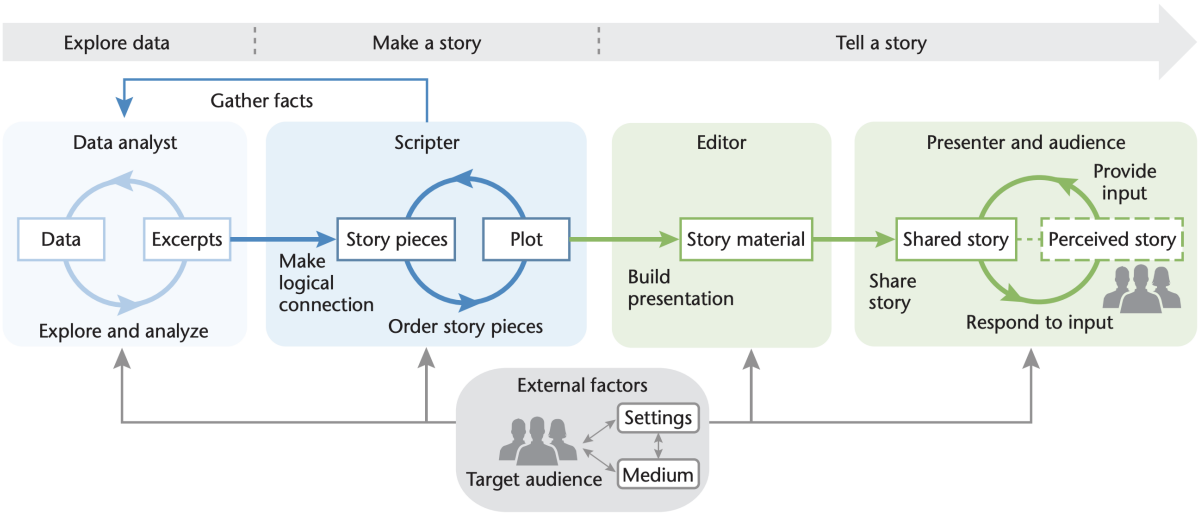


Figure 1.5: Storytelling process described by Lee et al. [2015].

To address RQ4, inspired by infographics, data comics, and cheat sheets for programming languages, we present Data Visualization Cheat Sheets for learning, teaching, and the regular use of visualization techniques (Chapter 6). We created cheat sheets by ourselves, and co-design with experts in a workshop iterated the designing and evaluated the cheat sheets with participants by doing tasks with the aid of our cheat sheets. To obtain first-hand feedback on the understandability and perceived usefulness of the cheat sheets, we conducted a qualitative user study with 11 participants reading visualizations unfamiliar to them with the aid of the cheat sheets.

RQ5) How to use data comics for promoting real-world communication, e.g., reporting controlled user studies?

Previous chapters provide initial knowledge about using and creating data comics, including the merits and drawbacks, creation workflow, and designing guidance. So, how well do data comics perform in solving real-world communication challenges? Scientific reports that include complex processes, analysis, and result presentation are data-driven communication in a way. Could data comics promote communication efficiency and engagement with scientific study reports?

We tried to use data comics to report controlled user studies in human-computer interaction, as it is the area where the author of this thesis is from. We elaborated a 10-stage framework to guide the storyboarding of data comics for controlled user studies. Then conducted collaborative design process and guidelines engaged participants in storyboarding comics for their studies (Chapter 7).

RQ6) What could interactions add to data comics?

So far, data comics encountered in the wild and research are mostly static. Interaction is

widely applied in narrative visualizations, which can enable viewers to consume a story in different ways, give more agency around how the narrative unfolds, and eventually may improve comprehension and recall in storytelling. However, how interaction changes how people read data comics is still unknown.

In chapter 8, we conducted a systematic review of traditional (non-data) interactive comics to understand which interactions the comic community is currently providing to support some of the six goals. By making interactive data comics by ourselves, we explore interactions that enable non-linear storytelling, personalization, levels of details, and explanations.

RQ7) How to create interactive data comics?

Creating data comics is a labor-intensive process. Making web-based interactive data comics requires proficiency in interaction design and front-end programming. However, existing prototyping tools such as Adobe XD and Figma could not sufficiently support the rapid creation of interactive data comics.

To facilitate understanding and creating interactive data comics, we present interactive operations and propose a lightweight scripting approach to add these operations to a set of existing comic panels (Chapter 8). We engaged in a two-week-long design process with six illustrators, designers, and postgraduate computer science students. Feedback suggests that our scripting approach is understandable by non-programmers and fosters exploration of the potential of interaction for data comics.

1.3 Thesis Statement

Data comics are an emerging genre for data-driven storytelling, leveraging the power of data visualization and visual storytelling with comics. Data comics contain mixed features of other visual storytelling genres, combining the linearity of videos and the freely exploratory features of infographics. However, few examples exist, and little is known about its merits, drawbacks, usage, and creation process. The author of this thesis presents *Data Comic Creation Model* (chapter 3) to set a theoretical framework for the knowledge of data comics, involving the essential elements (visualization, flow, narration, words and pictures [Bach et al., 2017a]), and its creation workflow. The thesis presents a further exploration of those elements, which sets the fundamental knowledge of this format, and suggests effective methods to promote data comic creation.

The research started with a hypothesis that data comics are more effective and engaging than infographics and magazine-style formats for presenting data stories. We compare the effectiveness of data comics with infographics and illustrated texts by analyzing quantitative and qualitative data collected from controlled studies in the lab and in the wild (Chapter 4). Data

comics with a higher level of text and image integration are seen as more engaging, helpful to stay focused, overall more engaging, and can improve understanding for most cases. With the overall positive results from the study, we aim to build a more systematic understanding and an effective workflow to guide data comic creation. We conduct hands-on data comic workshops in an undergraduate course with a focus on the stages from data collection, visualization, narrative creation, and drawing comics (Chapter 5). The structured plan and materials for data comic creation workshops provide reusable resources for future practice and evolve into a model that reflects a structured knowledge of data comics. The five layers within this model imply the important stages and elements of the data comic creation process. We further explore questions from the layers of the model. To support the creators with explaining data visualizations in data comics, we present "cheat sheets" for data visualization techniques (Chapter 6), in which we define six types of sheets for six common yet non-trivial visualization techniques and refine them through an iterative method. The design and resources of the cheat sheets are gathered from visualization lecturers, research publications, then evaluated with visualization novices. Visualization cheat sheets promote the power of data comics for presenting stories with complex visualizations. Besides visualization techniques, narrative plays a vital role in assembling individual pieces of information as a whole. So far, data comics have been used in the wild for public engagement and telling personal data stories, yet data-driven story with complex plots and information is a challenge. We engage data comics for scientific reporting as a complement to visual abstract. We collaborated with other authors of HCI papers to create a set of examples and concrete design solutions for communicating the common stages of HCI studies and statistical analysis (Chapter 7). The complexity of study procedure, data analysis process, and result explanation in scientific reports raise the demand for data filtering and information navigating. When encountering a data story, audiences have different demands of information, relations with the data, and desires to interact with the data. We get inspired from interactions from non-data comics and introduce interactive operations specified by a lightweight specification language for designers to add interaction to data comics (Chapter 8). In a creation workshop, participants use our approach to build interactive data comics and achieve multiple narrative purposes.

Data comics are an effective genre for data-driven storytelling, but creating data comics requires multiple skills. We hope this thesis sets a foundation for understanding and using data comics, fostering tool development, and making data comics more accessible to a wider audience and scenarios.

1.4 Outline

This thesis contains nine chapters. Related works in Chapter 2 provide a fundamental overview of data visualizations for communication, comics for data-driven storytelling, methodology, pedagogy, and tools for creating data-driven storytelling. Chapters 4, 5, 6, 7, and 8 are reproduced from the published papers with parts elaborated to serve the focus of this thesis (i.e., Motivation and Conclusion sections of these chapters). The original papers are completed with co-authors listed under the chapter titles in this section.

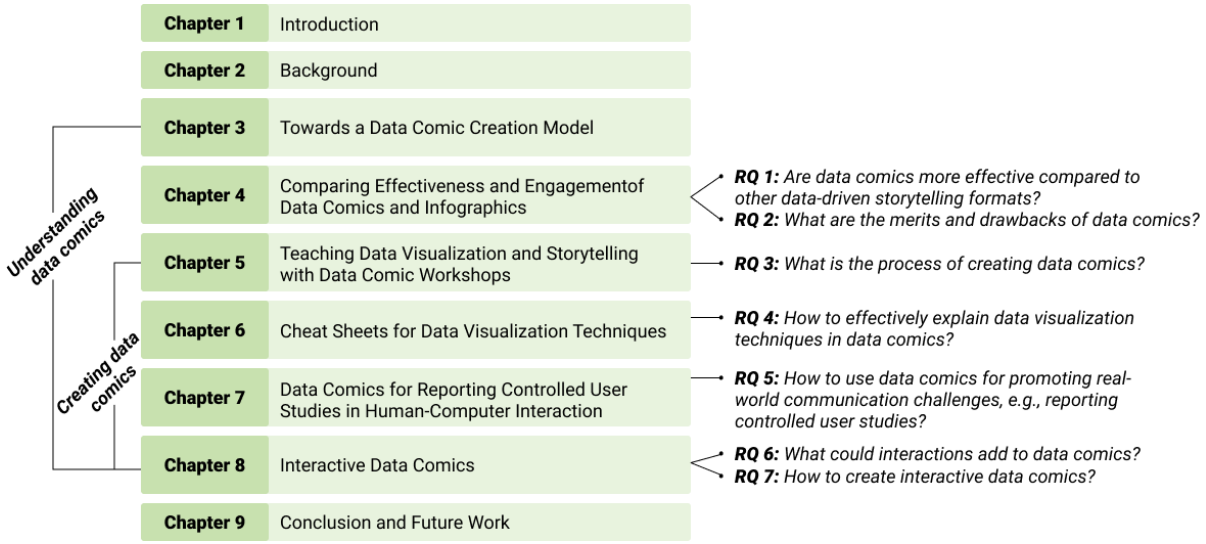


Figure 1.6: The outline of this thesis with research questions (RQ) and corresponding chapters.

Chapter 3: Towards a Data Comic Creation Model

Chapter 3 introduces the Data Comic Creation Model—a structured framework describing the elements of data comics, and a workflow for creating data comics. This model contains five layers that present a high-level overview of design consideration of data comic creation. The five layers from the bottom to top are *Communication Objective*, *Message*, *Narrative Structure and Explanatory Visualization*, *Visual-Narrative Integration*, and *Graphic Appearance*, the order reflects one possible process of data comic creation. The following chapters further elaborate on the form of structured layers and the essential elements within the model, including the using and explaining data visualizations techniques (Chapter 6), applying data comics for complex structured narrative (Chapter 7), and non-linear reading experience with interactive data comics (Chapter 8).

Chapter 4: Comparing Effectiveness and Engagement of Data Comics and Infographics

(published full paper: **Wang, Z.**, Wang, S., Farinella, M., Murray-Rust, D., Henry Riche, N., and Bach, B. (2019). Comparing Effectiveness and Engagement of Data Comics and Infograph-

ics. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.) My contribution to this study includes designing materials and questionnaires in the lab study, conducting the lab and in-the-wild studies, and being involved in analyzing data.

This chapter investigates the research questions of:

RQ1) Are data comics more effective compared to other data-driven storytelling formats?

RQ2) What are the merits and drawbacks of data comics?

A visual narrative consists of textual and visual contents as the basic channels, different formats primarily varied by how the elements of these two channels are ordered and presented at a time Segel and Heer [2010]. To present an understanding of how linear reading order with high integration of text and pictures affect the effectiveness of communication with data comics, this thesis starts with an empirical study that compares the effectiveness and engagement of data comics with other two popular static genres (i.e., infographics and text-plus-picture formats). This study presents a lab study and an in-the-wild study, showing qualitative, quantitative results and a reflection of design suggestions. The finding suggests the overall preference of data comics by participants and identifies its strength in explaining visualizations and presenting linear narratives compared to infographics and illustrated texts. The finding also helps to understand the respective roles of investigated formats and inform the design of more effective data comics. In particular, design guidance for the *Visual-Narrative Integration* layer, e.g., "balancing repetition and highlighting" and "balancing sequence and overview".

Chapter 5: Teaching Data Visualization and Storytelling with Data Comic Workshops

(Published extended abstract: **Wang, Z.**, Dingwall, H., and Bach, B. (2019). Teaching Data Visualization and Storytelling with Data Comic Workshops. In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems.) My contribution to this workshop includes taking part in preparing materials and participating in running the workshop and analyzing the data comics created by the participants together with another constructor.

This chapter investigates the research question of:

RQ3) What are the steps of creating data comics?

The promising results of using data comics for presenting data stories in Chapter 4 motivate us to move from understanding to creating data comics. Conducting creation workshops is an effective method to test methods and materials, also to discover challenges and solutions with participants. This chapter presents a data comic workshop with 23 illustration students. The methodology for the revised data comics workshop is informed by two previous workshops. We present our structured approach of conducting such a workshop, including slides, materials, a collection of data comics created by participants. Participants' works are promising, showing that they actively engage with the three main aspects of data comics: data, visualization, and

stories. This workshop identifies challenges such as creating a narrative, visualizing, and explaining data effectively motivates us to provide support for using and understanding visualization techniques for creators and audiences. We further explore solutions for these challenges in different stages of creation in the following chapters.

Chapter 6: Cheat Sheets for Data Visualization Techniques

(Published full paper: **Wang, Z.**, Sundin, L., Murray-Rust, D., and Bach, B. (2020). Cheat Sheets for Data Visualization Techniques. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.) My contribution to this workshop includes conducting expert workshops, expert interviews, iterating the design of cheat sheets for three types of visualization techniques, creating meta-cheat sheets for guiding one of the co-authors to create another three sets of cheat sheets, evaluating the readability in a study and conducting guideline workshops together with help from another constructor.

This chapter investigates the research question of:

RQ4) How to effectively explain data visualization techniques?

Data visualizations reveal insights from data and take an essential role in data comics to make messages comprehensible. From chapter 5, we learned that explaining visualizations to audiences within data comics is challenging for the creators. Chapter 6 explores how to properly decode visualizations to facilitate understanding of the visual mappings, graphical conventions, symbols, marks, layouts, and patterns. Inspired by the graphical explanation in infographics, comics, assembly instructions, and cheat sheets for programming languages, we present a set of cheat sheets for data visualization techniques. These cheat sheets are designed for facilitating understanding and using different types of data visualizations. Each cheat sheet has visual material covering the aspects vital for understanding a technique, such as anatomy, common patterns, or pitfalls. We refined them through an iterative method involving expert interviews, workshops, audience feedback, design-by-example, a readability study, and a guideline workshop. The visualization cheat sheets could be applied in workshop and classroom settings to help teach data visualization techniques and become an essential part of the data comic design repertoire.

Chapter 7: Data Comics for Reporting Controlled User Studies in Human-Computer Interaction

(Published full paper: **Wang, Z.**, Ritchie, J., Zhou, J., Chevalier, F., and Bach, B. (2020). Data Comics for Reporting Controlled User Studies in Human-Computer Interaction. IEEE Transactions on Visualization and Computer Graphics.) My contribution to this paper includes creating data comic examples, constructing the co-design session with the help of another constructor, and participating in abstracting the 10-stages of reporting together with other co-authors.

This chapter investigates the research question of:

RQ5) How to use data comics for reporting quantitative user study and data analysis?

The previous chapter provides support for explaining complex visualization techniques. Managing messages and complex narratives is another challenge with data comics, especially, organizing the information in an informative way meanwhile less overwhelming to the audience. This chapter investigates applying data comics with complex and structured linear narrative, i.e., reporting controlled studies in the domain of human-computer interaction (HCI). While many studies in HCI follow similar steps in explaining hypotheses, laying out a study design, and reporting results, many of these decisions are buried in blocks of dense scientific text. This chapter proposes leveraging data comics as study reports to provide an open and glanceable view of studies by tightly integrating text and images, illustrating design decisions and key insights visually, resulting in visual narratives that can be compelling to non-scientists and researchers alike. Use cases of data comics study reports range from illustrations for non-scientific audiences to graphical abstracts, study summaries, technical talks, textbooks, teaching, blogs, supplementary submission material, and inclusion in scientific articles. This chapter provides examples of data comics study reports alongside a graphical repertoire of examples, embedded in a framework of guidelines including what message could be reported by data comics. This framework is iterated upon and evaluated through a series of collaborative design sessions. Finally, we show that our method allows people to create clear and compelling visual summaries of complex research designs.

Chapter 8: Interactive Data Comics

(Published full paper: **Wang, Z.**, Romat, H., Chevalier, F., Henry Riche, N., Murray-Rust, D. and Bach, B. (2021). Interactive data comics. IEEE Transactions on Visualization and Computer Graphics.) My contribution to this research includes creating interactive data comic examples, taking part in abstracting interactive operations, conducting the workshop to test the design approach with another constructor.

This chapter investigates the research questions of:

RQ5) What could interactions add to data comics?

RQ6) How to create interactive data comics?

Chapter 4 suggests that data comics are effective for presenting overview and details at the same time, and guiding viewers through a linear reading sequence. However, the linear narrative structure shows a limitation in providing content for personal interests and enabling active exploration. Chapter 8 pushes data comics further by adding interactions. Inspired by data comic design patterns [Bach et al., 2018c], we present a set of *operations* tailored to support *data comics narrative goals* that go beyond the traditional linear, immutable storyline curated by the author. These include adding and removing panels into pre-defined layouts to support branching, changing perspectives, accessing detail-on-demand, providing and modifying data,

and interacting data representation to support personalization and reader-defined data focus. We propose a lightweight specification language for designers to add interactivities to static comics, while comic assets are created using traditional graphics editing tools or digitized hand-drawn sketches. To assess the viability of our authoring process, we recruited six professional illustrators, designers, and data comics enthusiasts and ask them to craft an interactive comic over two weeks, allowing us to understand the authoring workflow and potential of our approach. Result suggests that our scripting approach is understandable by non-programmers and fosters exploration of the potential for interactive storytelling.

Chapter 9: Conclusion and Future Work.

Are data comics an effective genre for data-driven storytelling? How to create effective data comics? The last chapter summarizes the whole thesis by highlighting the contribution of each chapter, stating how individual chapters answer the seven research questions in Chapter 1. This chapter is finished with planned future work that can be seen as short or mid-term further research including an evaluation and reflection to the conceptual data comic creation model; a study of styles in data comics; providing an automatic or semi-automatic tool for rapid data comic creation; extending the scenarios and applications of data comics.

1.5 Contributions

The contributions of this thesis include:

- The author of this thesis introduces the Data Comic Creation Model to reflect the contribution of this thesis and the working process of creating data comics (Chapter 3).
- Qualitative and quantitative results of a controlled lab study and in-the-wild study compare recall, comprehension, and engagement of illustrated text (magazine style), infographics, and data comic. Reusable study resources study materials, combined lab and in-the-wild study methods, and questionnaires for future research (Chapter 4).
- The results fill the knowledge gap of the strength and drawbacks of data comics, e.g., clear sequencing increases the readers' ability to focus and navigate spatial-temporal information, while panels help divide information into easily memorable chunks, with rows grouping individual messages into higher-level messages. On the contrary, not highlighting important information in comics, such as visual encoding, can lead to participants overlooking details. These results suggest that increased text-picture integration and more reading guidance can lead to better understanding (Chapter 4).

- Qualitative results lead to guidance for designing effective data comics (Chapter 4).
- We provide a guideline of steps, activities, and materials for conducting data comic creating workshops `https://datacomics.github.io/` (Chapter 5).
- We collect data comics created by our participants in data comic creation workshops and examples created by others at `https://datacomics.github.io/` (Chapter 5).
- We present cheat sheets for data visualization techniques, including six types of cheat sheets covering different aspects of each data visualization at `https://visualizationcheatsheets.github.io/` (Chapter 6).
- We present a set of meta-cheat sheets as guidance for designing each type of cheat sheet in the future `https://visualizationcheatsheets.github.io/` (Chapter 6).
- We contribute a set of examples of how comics support that reporting controlled studies `https://statscomics.github.io/` (Chapter 7).
- We provide an initial set of concrete ideas for communicating common stages of HCI studies and statistical analysis and may serve as a starting point for future exploration. We show that our new method allows people to create clear and compelling visual summaries of complex research designs (Chapter 7).
- We present examples of interactive comics in a gallery. This initial step towards understanding the design space of interactive comics can inform the design of bespoke tools and experiences for interactive storytelling `https://interactivedatacomics.github.io/` (Chapter 8).
- We contribute a tool for constructing interactive data comics with a concise scripting language called COMICSCRIPT, a set of design operations for adding interactivity in data comics, e.g., inputting the value, arranging, and manipulating layout (Chapter 8).
- We provide self-contained tutorials for using the tool and COMICSCRIPT to make interactive data comics (Chapter 8).

Chapter 2

Background

Data comics touch multiple disciplines such as data visualization, data-driven storytelling, human-centered design, comics, creative design, and education. This chapter covers the literature based on the notion of *what data visualization and data-driven storytelling are*, *why comics can be used for telling data stories*, and *how visual data stories are created*. This chapter begins with an overview of data visualizations and how they could help understand messages from data. By looking at the factors that affect communication efficiency with visualization, related works highlight the strength of narration in engaging the audience to understand data in context. As one of the storytelling formats, section 2.3 introduces typical characteristics of comics that imply the potential of presenting data-driven stories. Section 2.5 outlines existing design methodologies (i.e., design models, workshops, and tools) and an overview of relevant systems for creating data stories and comics.

2.1 Overview of Data Visualization

Abundant information is collected and analyzed through data, to explore the world and make sensible decisions, people need to be able to interpret the information from data. To make data more natural to be comprehended by the human mind, *data visualizations* shape the way people understand data by "constructing a visual image in the mind" [Little, 1956], they use "interactive, visual representations of data to amplify cognition" [Card, 1999]. Visualizations are presented in many different ways, which can be spatial using maps, temporal relation with a timeline, a basic pie chart, a bar chart, or a line chart generated in Excel, or even scaling representations with physical items (i.e., data physicalizations) [Dragicevic et al., 2020]. Basically, data visualizations are built with basic graphic elements such as lines, points, and areas, with various properties like colors and textures, or with icons and symbols [Bertin et al., 1983]. Encoding data with

visual variables makes trends, patterns, and outliers within large data set more observable, so that people could discover, compare, and explain the knowledge transformed from data.

Data visualization is used in various scenarios and disciplines, ranging from medical and engineering domains to financial markets and social science [Johnson et al., 2005], and become popular in the field of journalism [Valero et al., 2014]. For example, data visualizations in academic publications (e.g., Figure 2.1 (a)) are used to report the data collected in experiments, which require the data to be precisely presented; support lay audience to understand data and insights about finance, climate, and sociology in news articles, to demonstrate multiple aspects of an event or a topic, and be more communicative with annotations and explanations through a linear narrative (Figure 2.1 (b)); share personal data in a creative way by visualizing self-collected data on a postcard-sized sheet of paper (Figure 2.1 (c)).

Figure 2.1: Examples of data visualizations used in different scenarios: (a) Results for time, error, and subjective measurement reporting in academic publications by Bach et al. [2018a]; (b) Data visualizations in news article [Economist]; (c) Hand-drawn data visualization on a post card [Lupi and Posavec, 2016].

According to the purpose, data visualizations are categorized by exploratory and explanatory visualizations [Iliinsky and Steele, 2011]. Exploratory data visualizations are particularly used to explore unknown relationships within data, namely, "looking at data to see what it seems to say" through obtaining, cleaning, profiling, analyzing, and interpreting data [Tukey et al., 1977]. While explanatory data visualizations present results of the data analysis and help the viewer to understand the statistical data by illustrating and explaining messages. In practice, although analysis and communication normally happen separately, the boundary between visualizations created for exploration and explanation can be blurry and overlapped (Figure 2.2). Communication may take place in both the exploratory stage and vice versa, for instance, the analysis

of a computational notebook needs to be clearly explained for the analyst's future self, collaborators, or the manager [Rule et al., 2018]. Viegas and Wattenberg [2006] introduce the concept of communication-minded visualization (CMV), emphasizing the role of visualization-based communication in collaborative analysis and social interactions, suggesting communication and social activities surrounding data analysis can be either synchronous and asynchronous. This thesis focus on improving the effectiveness and engagement of visualizations for communicating with general audiences, including both visualization experts and novices (Figure 2.2).

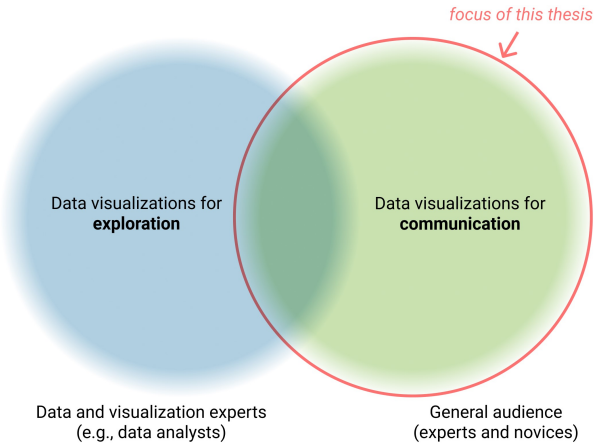


Figure 2.2: According to the purpose, data visualizations can be categorized as for exploring data used by data and visualization experts and communicating the insight from data to general audiences ranging from novices to experts. This thesis investigates creating data comics to promote communication with data visualizations.

The effectiveness of communication with data visualizations is defined by multiple factors. Comprehension and memorability are commonly measured in studies [Borkin et al., 2013, Anderson et al., 2011], while Kennedy et al. [2016] argue that accessing specific information and quickly remembering messages are not always the goals of using visualizations, they extend the definition of effectiveness with wider dimensions, such as stimulating emotions, provoking interest and questions. That is, the effectiveness varies among audiences, Shah and Hoeffner [2002] categorizes three major factors that influence the engagement when people encounter a visualization: the visual characteristics of visualization, an audience's knowledge about visualizations, and an audience's knowledge and expectations about the content of the data in a visualization. In data visualizations, content is engineered in a way that conforms to human perception, visual features such as color, size, and orientation are used to direct the attention of viewers [Treisman and Gelade, 1980, Haroz and Whitney, 2012]. To accompany the visuals, textual components including titles, labels, annotations, explanations, and comments take a significant role in promoting understanding [Sorapure, 2019]. Visualization creators also apply allusions, metaphors,

and cultural references to convey messages [Ziemkiewicz and Kosara, 2009]. Furthermore, Kennedy et al. [2016] identified six factors that affect the engagement with data visualizations: whether audiences interested in the subject matter; whether the audience trust the source of a visualization; whether a visualization fits the audiences' beliefs and opinions; whether an audience wants to commit time on reading a visualization; the emotions of first impressions; the confidence and skills the audience owned for decoding visualizations. The factors above reflect challenges of drawing attention from the audience before and during the communication process, and on the other hand, imply the direction of efficient visualization creation.

2.1.1 Data Visualization Techniques and Data Visualization Literacy

While visualization techniques are intended to make information comprehensible, many techniques require some degree of learning and understanding of the visual mappings, graphical conventions, symbols, marks, layouts, and patterns in order to properly decode the representation [Börner et al., 2016, Lee et al., 2016, Maltese et al., 2015]. Visualization techniques are helpless for understanding data if the audiences are not familiar with the visualization technique. Similar to understanding words before being able to read articles, interpreting the visual representation and transforming visual patterns that can be read and understood by people are central to the practice of data visualization.

Many forms of literacy are considered essential in a data society, such as data literacy, numeracy, and visual literacy [Bristor and Drake, 1994, Schield, 2004]. *Visualisation literacy* is an essential factor in communication with data stories because the audiences of data stories are broad, and visualization techniques, especially the uncommon ones (e.g., parallel coordinates and adjacency matrices), are not familiar to lay audiences. [Boy et al., 2014] provide a formal description for visualization literacy as "the ability to confidently use a given data visualization to translate questions specified in the data domain, as well as interpreting visual patterns in the visual domain as properties in the data domain". This definition emphasizes the ability to translate information from the specific application domain to the visual domain and back.

The translation requires correctly interpreting the graphical conventions of visualization in order to create a consistent mapping between data and graphics so that higher-level domain tasks can be mapped onto lower-level visualization tasks. Consequently, Börner et al. [2016] define visualization literacy at a task level including "interpreting patterns, trends, and correlations in visual representations of data". Bertin et al. [1983] describes three sequential stages of interpreting a visualization: an *external identification stage* where a reader recognizes the frame of visual encoding (e.g. axes, labels and dimensions); an *internal identification stage* where a reader

identifies visual characteristics (visual patterns); and a final *perception of correspondence stage* where a reader analyzes the content and extracts messages from the visualization. Similarly, the NOVIS Model [Lee et al., 2016], describes readers encountering unfamiliar visualizations as wandering through the following five activities: *encountering*, *constructing a frame*, *exploring the visualization*, *questioning the frame* and *floundering on the visualization*.

Materials to support the learning of visualization techniques are presented in various formats with different scenarios in mind. Scientific papers introducing a novel technique can often be considered a primary source of learning about a technique. To this end, many papers include conceptual descriptions of the technique, diagrams on the creation of examples and demonstrate visual patterns (e.g. [Henry and Fekete, 2006, Heer et al., 2009, Bach et al., 2014b, 2016b, 2017b]), although some focus exclusively on the visual patterns (e.g. [Haroz et al., 2016, Wilkinson et al., 2005, Behrisch et al., 2016]). Textbooks (e.g. [Kirk, 2016, Cairo, 2016 - 2016, Rendgen et al., 2012]) tend to offer summaries of techniques alongside comprehensive knowledge about visualization design principles, goals, workflows, visualization categories, encoding, collections of examples, and specific visualizations techniques. For example, Munzner [2014] explains visualization techniques on a by-case basis, showing example implementations and discussing advantages and drawbacks. Researchers also propose the concept of learning unfamiliar visualizations by analogy [Ruchikachorn and Mueller, 2015], and tool for progressively constructing visualizations from visual channels and units [Wang et al., 2019a].

More systematic descriptions of visualizations are found in Wilkinson [Wilkinson, 2012], who demonstrates a catalog with mathematical principles and fundamental components of quantitative statistics graphics. However, to quickly understand and use visualization techniques, more succinct and novice-friendly instructions are required. Kirk [2016] systematically presents 39 visualization techniques, each being presented on a single page, with primarily textual descriptions, but with consistently structured information covering *description*, *examples*, *how to read it*, *presentation tips* and *variations & alternatives*. Similarly, some online resources gather information about very specific techniques, prepared by different educators and without general structure, e.g, on WikiPedia [Wik] and online videos [Herke, 2019], or collections of pitfalls, but without graphical explanations [L. Satterthwaite et al., 2019]. The DataVizProject [Ferdio, 2019] and Chart doctor [Smith et al., 2019] list types of visualization techniques with examples, categorization, and descriptions, and backed up with abstract pictograms, but do not provide specific aid for understanding and learning. The catalog lists 60 visualization techniques, each of which comes with an example, a textual description, anatomy, links to tools, and a loose tag list (e.g., "data-over-time", "patterns") [Ribecca, 2019]. For some examples, it lists similar techniques to aid selection. With the same goal of identifying suitable chart types, *data-to-viz.com* [Healy,

2019] provides similar contents yet adds “common mistakes” and “build your own” with R and python galleries. While the collection covers many techniques, many explanations lack detail, depicted patterns are rarely and only superficially explained, and there is no deeper support in engaging with learning and using the technique.

These materials provide rich resources for understanding and using data visualization techniques. However, two obstacles may impair the effectiveness of communication with data visualizations. First, huge learning cost is required for learners to understand scientific papers or lengthy textual explanations of books. Second, much of this material talks about general topics in visualization such as deception, design, presentation, storytelling, or layouts. For audiences, understanding specific visualization techniques encountered in data stories is still a challenge. This thesis provides solutions inspired by the visual language of comics to support lay audiences for understanding visualizations while reading.

2.2 Visual Data Stories

Besides interpreting visualization techniques, delivering visualizations in a compelling way to catch the audience’s attention is highly desired. Storytelling is an ancient art of communication and used in many different fields. Explaining data under context with transitions between events, storytelling is also a popular concept in communicating insight from data. Segel and Heer [2010] propose the term “narrative visualization” to describe the storytelling with data-driven graphics. Distinct from visualizations for exploration, Kosara and Mackinlay [2013] name storytelling affordances to describe “the feature of a visualization that provides a narrative structure and guides the reader through a story”. This thesis follows the definition of data-driven stories by Riche et al. [2018] — “Data-driven stories are stories that are data-driven in that they start from a narrative that either based on or contains data and incorporate this data evidence, often portrayed by data graphics, data visualizations or data dynamics to confirm or augment a data story”. The definition highlights the features, i.e., data evidence and data visualizations, that distinguish data-driven stories from general stories.

How data-based evidence could be presented with compelling narratives is a highly-focused question in data visualization [Tong et al., 2018]. In general, the graphic of visualizations would hardly stand alone to deliver understandable and compelling stories. Thus information is remembered better when it is supported with pictures [Levie and Lentz, 1982], even more so when presented through both visual (e.g., diagrams, illustrations, graphs, or animations) and textual (written or spoken) channels at the same time [Paivio, 1990, Clark and Paivio, 1991, Baddeley, 1997, Sultanum et al., 2018]. Moreover, data visualizations have been found less memorable than natu-

ral scenes [Borkin et al., 2013], yet adding embellishments and unique presentations can improve memorability [Bateman et al., 2010, Borkin et al., 2013]. Studies comparing comics and other visual formats [Kraft et al., 2016] with text-only material [Aleixo and Sumner, 2017] and illustrated texts [Mayer and Gallini, 1990] confirm these trends and show increased memorability. Other studies have found that in cases where text is extraneous, closely integrating text and picture can distract the reader and hinder learning [Chandler and Sweller, 1991, Rop et al., 2018], while providing inappropriate graphics can impede understanding [Lin and Chen, 2007]. Integrating pictures and text in different ways leads to a different style of reading and perceiving a data story.

2.2.1 Genres in Data-Driven Storytelling

The term *"genre"* has been coined by Segel and Heer [2010] to describe types of narrative visualizations: magazine style, annotated chart, infographic, video, slide-show, flowchart, positioned poster, and comics (Figure 2.3). To understand the common features and reading styles of these genres, some of them can be classified as *space-oriented* (infographics, posters, annotated charts) and others as *time-oriented* (videos, slide-shows). Space-oriented genres present and communicate through space: layout, size, visual and graphical elements organized on a canvas, equally and instantaneously accessible to the reader, leaving it to the reader which elements to watch, how long, and in which order. The time-oriented genres present and communicate linearly over time; they have an order with a fixed beginning, a middle, and an end; a development over time, an argumentation, and they can present change over time through animations [Amini et al., 2015, Chevalier et al., 2016]. Observers are locked in the order and pace at which they perceive the content. Research has shown that narrative structure is intrinsically easier to remember and facilitates readers' engagement and persuasion [Green et al., 2003]. However, this requires the audience to remember, to continuously pay attention, and to adapt to the pace chosen by the presenter. In short, while animated explanations can be useful to present dynamic content [Höffler and Leutner, 2007], animations often risk showing "too much, too fast" [TVERSKY et al., 2002].

The third group of genres can be described as both spatial *and* temporal. This includes magazine-style (text with figures) as well as comics. Comic scholar Groensteen writes that "[c]omics is not only the art of fragments, of scattering, of distribution; it is also an art of conjunction, of linking together [Groensteen, 2007, p.22]. Comics combine linear narration with non-linear figures and, as Eisner notes, the "partnership of words with imagery becomes the local permutation. The resulting configuration is called comics and it fills a gap between print and film." [Eisner, 2008, p.xvii]. So, how could comics be both spatial and temporal, and how does this unique feature support communicating? Understanding the essential working

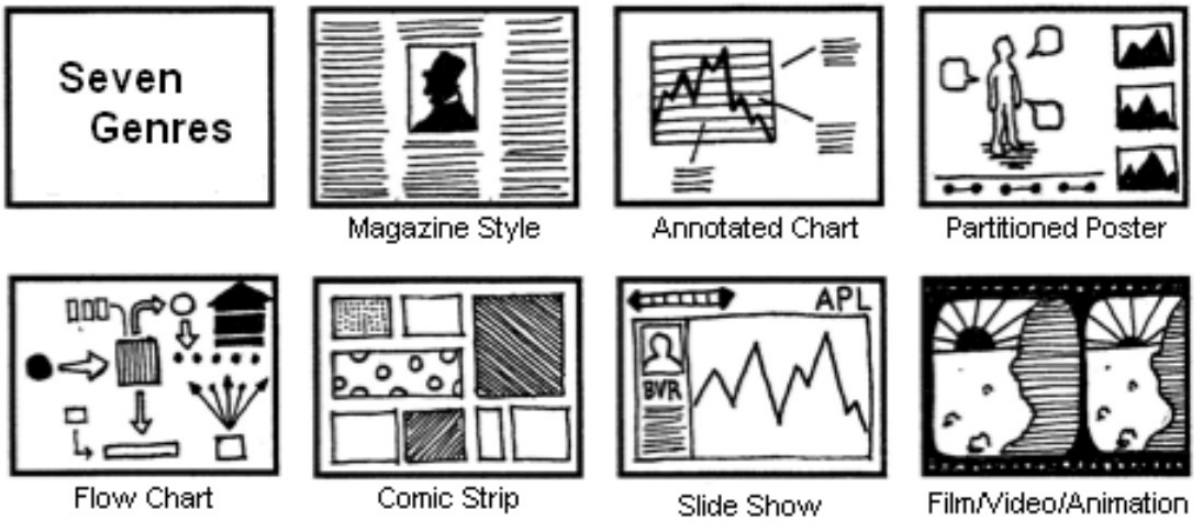


Figure 2.3: The seven genres for presenting narrative visualization [Segel and Heer, 2010]

mechanism of comics can inform the use of comics for presenting data-driven stories.

2.3 Comics for Communication

According to McCloud [1993], comics are defined as "Juxtaposed pictorial and other images in deliberate sequence, intended to convey information and/or to produce an aesthetic response in the viewer". Under this definition, sequential arts such as Bayeux Tapestry and Egyptian Hieroglyphics have a long history for storytelling and communication [McCloud, 1993]. In addition to placing images in sequential order, comics offer a set of unique characteristics for communication and optimal understanding, being highly accessible to a large audience, compatible with many different media, do not require a presenter, and can be read at one's own pace [Yang, 2008]. More than the cartoonish style, jokes, and superhero stories, comics are also popular for explaining science [Tatalovic, 2009, Farinella, 2018]. For scientific communication, comics can blend explanation (e.g. schemata, illustrations, data visualization) with narration, characters, and dialogue. While how could a comic encapsulates a specific *message* (or information) represented as an integrated combination of text and picture and build a deeper understanding in the audiences' mind? This section demonstrates characteristic features of comics — *panel*, *transition*, *layout*, *visual morphology* and how these features may benefit communication in 2D space.

2.3.1 Panel — A Window on a Visual Scene

The central narrative device in a comic is the *panel*, defined as a frame "depicting a frozen moment" [wik, 2017] and the camera through which the reader perceives the "world" [McCloud, 2011]. A single panel acquires meaning from its individual content as well as from its position within the global visual/narrative structure—a process defined as "braiding" (*tressage*) [Groensteen, 2007]. In fact, the notion of panels can be extended to the other narrative genres such as infographics, posters, slideshows, and videos. In infographics and posters (the space-oriented genres), a panel corresponds be any graphical component that can be understood as a narrative element and that is somehow separated from the rest of the visual space (e.g. through a border or empty space). For example, a panel can be a paragraph of text, a single image, or any combination of both. Sometimes, panels are made explicit by a drawn frame; sometimes, panels are separated from each other through white space (or even pages) and can be nested. In videos and presentations (the time-oriented genres), a panel corresponds to a single frame or a short scene, including the spoken or written word. Panels in videos and presentations are separated from each other through time. While panels in infographics are juxtaposed in space, panels in videos are sequenced into a unique temporal order. Transitions between panels can be abrupt (cut) or gradual using blurring techniques, or animated transitions [Heer and Robertson, 2007, Bach et al., 2014a].

2.3.2 Transition and Layout — Building Connections

To build the narration between individual panels, comics use transitions [McCloud, 1993] for preserving visual permanence [Yang, 2008] and allowing readers to read the story at their own pace. Relations between just two panels have been described as transitions by McCloud [1993] in comics, including time steps (*moment-to-moment*), actions (*action-to-action*), characters or objects (*subject-to-subject*), places or scenes (*scene-to-scene*), aspects (*aspect-to-aspect*), or no specific relation at all (*non-sequitur*) (Figure 2.4). Similar transitions have been described for data comics [Bach et al., 2017a] and which reflect the more general taxonomy by Hullman et al. [2013b] on sequences in narrative visualizations: temporal, spatial, granular, comparison, causal, dialog. These transitions and categories have been inspirational to our dimension of *content relation* with some categories maintained: temporal, granular, facets. Besides their linear structure, comics have a rich history of integrating unexpected logic into their structure, such as hierarchical linguistic grammars [Cohn, 2013b], repetition [Shores, 2016], or musical elements that carry temporality, emotion, and a joyous disregard for staying in their expected place [Brown, 2013]. Beyond the comics with a strong (linear) narrative, many artists have explored non-linear narrations (e.g., Figure 2.5), and the visual language by which time, space,

and narration can be expressed [Molotiu, 2009].

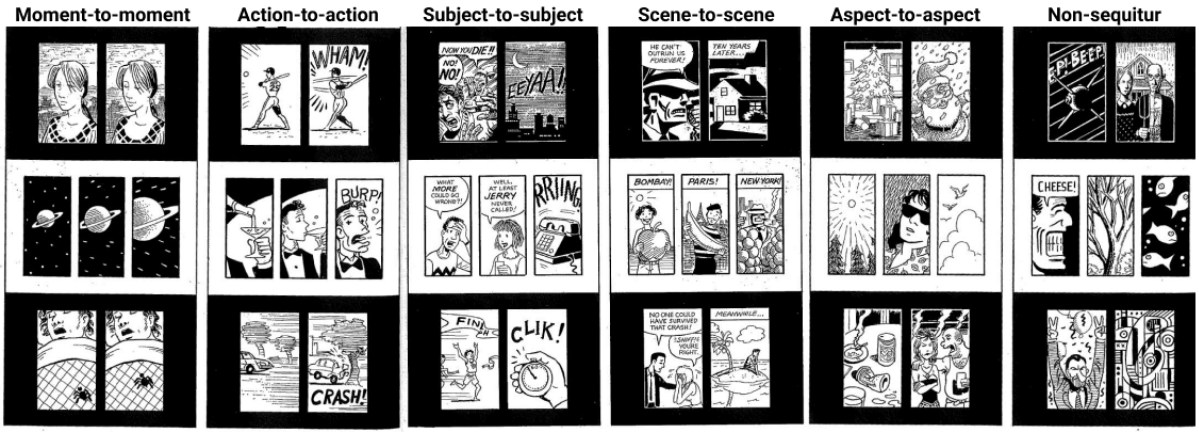


Figure 2.4: Six types of transitions between panels described by McCloud [1993].

Figure 2.5: Inventive uses of panel transitions: (a) Directing and redirecting to a panel to demonstrate an reappearance [Rabin, 2012]; (b) Using nested panels to present multiple points in time within a single location by [McGuire, 1989]; (c) Showing as character's daily routine with non-linear panels around the same location [Ware et al., 2012].

For storytelling in general, the use of sequence has been found to increase recall, facilitated by information being split into chunks [Black and Bower, 1979]. Psychological studies have investigated the effectiveness of the panel layouts [Cohn, 2013b, 2014] used in traditional comics through eye-tracking [Foulsham et al., 2016]. Additionally, the comic page layout allows a reader to consider information on multiple levels: a detailed explanation in sequence, as well as browsing the concepts and explanations simultaneously [Sousanis, 2015].

2.3.3 Visual Morphology — Articulation of Perception

Comics have a huge corpus of visual morphology. Morphemes are the smallest units of meaning in a language. In a similar manner, the visual language of comics uses morphemes to express meaning as well. for example, comics have the minimalist design style of conventional drawing to illustrate characters and objects (e.g.,xkcd comics [Munroe, 2015]), which is not complex to draw but successfully convey its message. Conventional drawing is an effective approach to demonstrate the context for a data story. In comics, texts are presented with different types of carriers to present various effects, such as a speech balloon for dialogue, a jagged-edge carrier for a loud shout, and a cloud-like bubble for thought. Indexical lines are another common morpheme in many surface forms. For instance, comic creators use arrows for pointing at something, speed lines, and action lines for motions radial lines for a light. The rich visual morphology in comics enriches the audiences' sensory experience by inspiring their imagination, and therefore remedy the limitation of the presentation in static and 2-dimensional interfaces [Cohn, 2013b].

With the powerful and expressive ability for communication, comics have been valued for their ability to promote public engagement [Negrete and Lartigue, 2004], e.g., in stem cell research [Amaral et al., 2015], nanotechnology [Lin et al., 2015], and many other domains [Farinella, 2018]. Comics have also been used for classroom instruction of STEM topics, showing promising increases in student engagement [Hosler and Boomer, 2011, Spiegel et al., 2013]. Comics have been used to teach concepts in statistics Takahashi et al. [2008], Gonick and Smith [1993], Peck et al. [2011] such as means, medians, and regression. The approach presented in this thesis differs from such uses of comics to teach statistics, yet uses the comic to promote comprehension and engagement with data-driven storytelling.

2.3.4 Data Comics

Inspired by comics, data comics aim to communicate insights in data through visualizations. Data comics support data-driven storytelling by making use of peoples' familiarity in reading and understanding comics along with the particular qualities of the medium [Zhao et al., 2015b, Bach et al., 2017a]. Zhao et al. [2015b] define data comics as "juxtaposing multiple visualizations into comic strip layouts consisting of a sequence of panels, each appropriately annotated and decorated with both visual and textual elements, and arranged into a sequence that progressively develops the overarching story told in the comic". Data comics combine elements from different genres, presented in a unique visual and narrative framework. Bach et al. [2017a] describe four essential components that help discussing the uniqueness of data comics: the presence of *data visualization*; *flow*, whether linear or non-linear; *narration* to include and illustrate contextual

information and create a compelling narrative as opposed to a factual presentation of charts; and *words-and-pictures* combined visually using multiple strategies.

While there are many examples of data stories that include some elements of comics (juxtaposition, a combination of text and image, etc), by the time this PhD project started, very few examples exist that combine these components into fully-fledged data comic. For example, Nature journal provides a high-quality example on the challenges of climate change [Monastersky and Sousanis, 2015] in 2015 (Figure 1.1), combining data visualization, narration, linear-and-non linear flow, into a compelling, visually striking, and even emotional piece. Cisneros [2011] created an interactive comic using Tableau Graphics. Additional examples include a scientific report on the change of water temperatures [Bernaerts, 2006], graph comics [Bach et al., 2016a] and examples in Bach et al. [2018c] (e.g., Figure 2.6). By exploring the dimensions of *flow* and *narration*, we presented data comic design patterns [Bach et al., 2018c] before this PhD, each pattern describing a set of panels with a specific narrative purpose (e.g., Figure 2.7), that allow for rapid story-boarding of data comics while showcasing their expressive potential. The description of design patterns implies a design space of data comics, which combines a dimension for spatial panel-layouts with a dimension for content relation between panels. However, creating data comics is still a difficult process, in part because of its interdisciplinary nature, and also because they lack tool supports, many manual works could take significant time and effort during the creation process. This thesis explores the creation process of data comics, aims to provide creation methods, reusable design solutions, and to set fundamental resources of knowledge and design guidance to foster the use and creation of effective data comics.

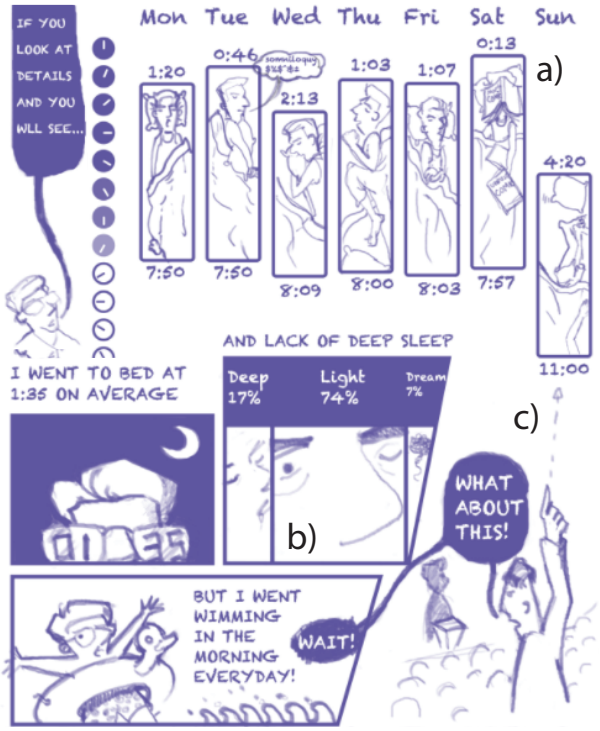


Figure 2.6: Data comic example employing design patterns: a) *temporal sequence*, b) *bar-chart panel*, and c) *flashback*. (©Zezhong Wang)

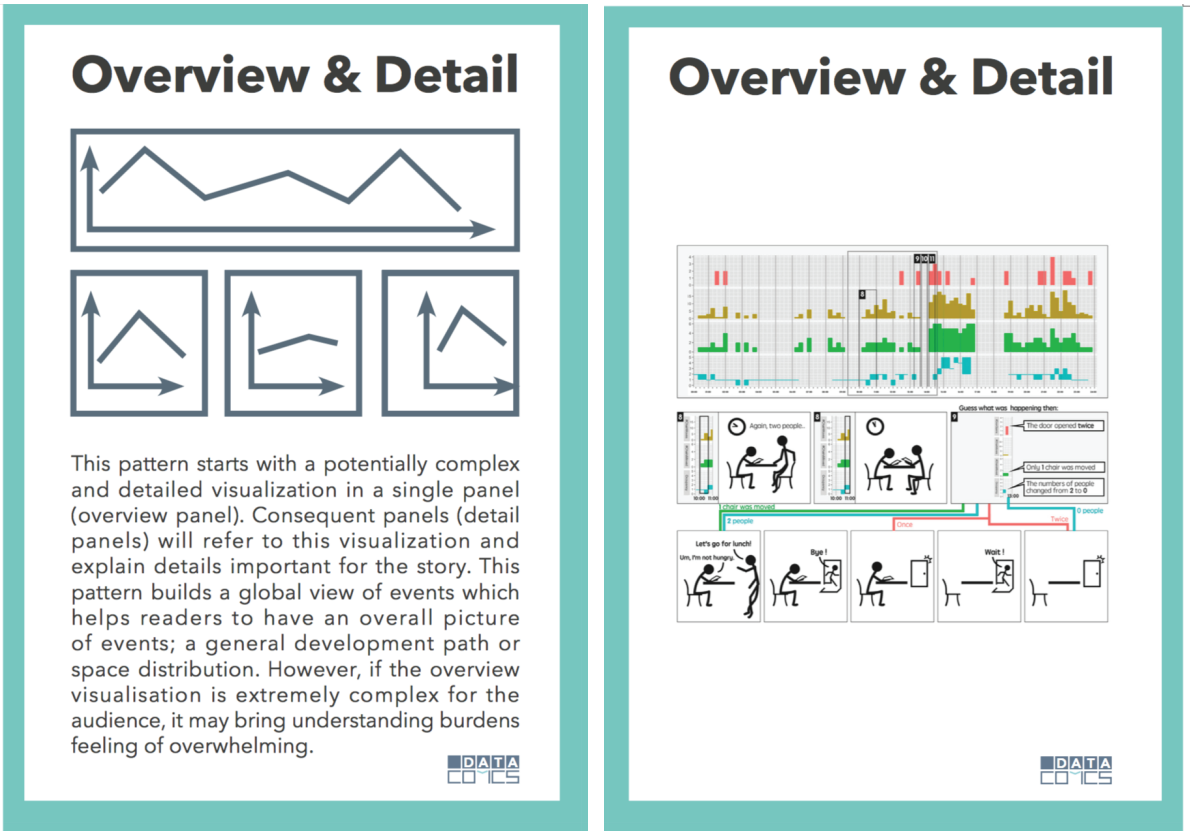


Figure 2.7: Example of a design pattern card for workshop, printed on A6 format: front (left) and back (right).

2.4 Methodology and Pedagogy for Creating Data-Driven Storytelling with Data Visualizations

In contrast to traditional comics, there is little accepted practical knowledge for creating data comics. Traditional comics employ steps such as story, narrative, characters, layout, and visual style, while data comics involve additional steps common to information visualization: data access, mining, stats, visual mapping, and explaining. Creating data comics is a complex process, more than assembling texts and pictures into panels, it implies careful engineering of the four essential components of data comics into one piece, and also creative-thinking, human-centered design thinking, be informative in data presentation. This section looks into creating methodology and pedagogy of designing information products and data-driven storytelling with data visualizations.

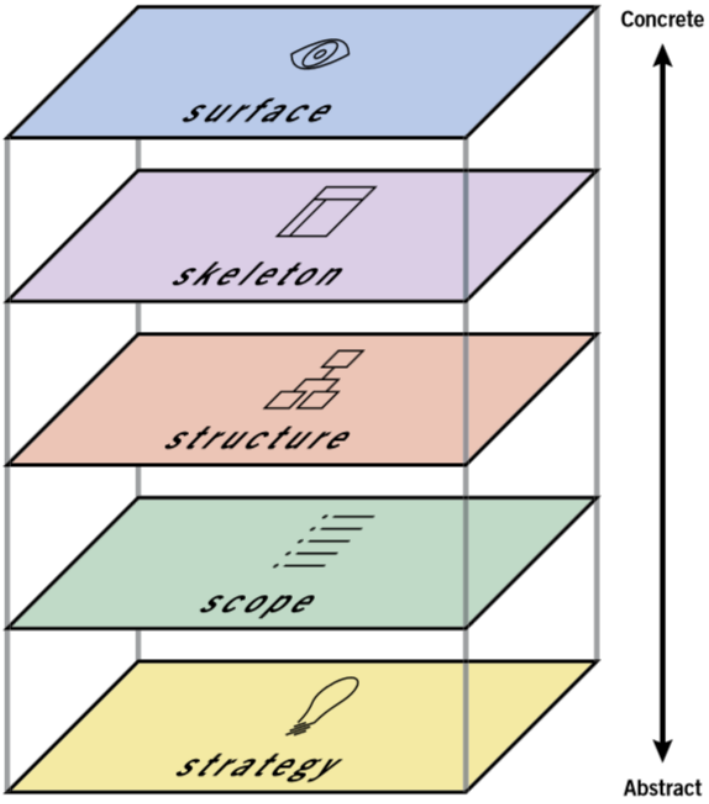


Figure 2.8: The Five Elements of User Experience Design by Garrett [2010]. There are five dependent layers, each level builds on the level before it, and they start with abstract level towards concrete level (from bottom to top).

2.4.1 Process for Creating Data Stories and Visualizations

Many academics and companies have in the past tried to develop models for design process [Council, 2007]. These models provide a high-level view of the design process from discovery problems or opportunities to design solutions. For example, a model presented by McDermott and O'Connor [2002] reflects a combination of the corporate design process and the individual designer's process. Cross [1989] introduces theories of how design and engineering work in partnership. The "Double Diamond" [Council, 2015] presented by Design Council divides the product design process into four phases (i.e., discover, define, develop and deliver); the five-plane model [Garrett, 2010] focus on the user-centered design process of functional and information products, the model includes five dependent layers from bottom-up indicates "strategy, scope, structure, skeleton, and surface" (Figure 2.8). Regarding a data story as an information-oriented product crafted for a target audience with a certain communication objective, the methodology of designing products (e.g., five-plane model [Garrett, 2010]) can inform the design process of data story creation.

Moreover, the data-visualization community has likewise several models that help developers consider and build visualization tools. For example, the nested model with four levels provides prescriptive guidance for determining appropriate evaluation approaches by identifying threats of each level [Munzner, 2009]. The four overlappings, multi-linear activities (i.e., understand, ideate, make and deploy) compose the common visualization design activities [McKenna et al., 2014]. The closest model that informs the data comic creation process is the visual data storytelling process presented by Lee et al. [2015], which presents three main components, i.e., "explore data, make a story and tell a story" (Figure 1.5). In particular, Lee et al. [2015] list activities of constructing a storyline, e.g., ordering, building logic connections, developing flow, formulating a message, and creating the denouement. However, this model requires adaptation to data comic creation by including interpreting visualizations and presenting stories in comic formats. Informed by the models mentioned above and the research in the following chapters, Chapter 3 contributes a model to demonstrate the process of creating data comics.

Models can provide a high-level overview of a general creation process, and illustrate the relationship among factors therein, to provide more practical guidance with detailed activities of creating data comics, this thesis includes conducting workshops as an important method to create data comics, discover questions, and evaluate creating methodologies.

2.4.2 Workshops and Activities

Workshops offer a unique opportunity to provide for active learning [Hearst, 2016] and engage participants through hands-on exercises and feedback. With project-based learning students work on challenges, resembling authentic real-world problems [Bell, 2010]. Consequently, workshops have become an integral part of any teaching in data visualization.

Workshops for creating data visualization usually involve types of data and visual variables, encoding principles [Cairo, 2013 - 2013, Lupi and Posavec, 2018], design and ideation methods [Huron et al., 2016, Roberts et al., 2016], interdisciplinary collaboration [Kerzner et al., 2019] and long-term teaching [Nolan and Perrett, 2016]. Procedure and materials are available for supporting data visualization creation in the classroom and workshop settings [Huron et al., 2016, He and Adar, 2017, KUN et al., 2018, Bach et al., 2018c]. For example, worksheets provide a structured guideline for conducting workshops, Roberts et al. [2016] suggests Five Design-Sheets indicating design by-sketching process. On the first sheet, creators begin with thinking and sketching possible outcomes, followed by creating three alternative designs on the second, third, and fourth sheet, and finally on the fifth sheet to select one of the design alternatives to add more details. Five Design-Sheets focus on creating interactive computer

systems and helping creators to think through different ideas and selecting the best design for the task. For supporting novice designers to create data visualizations, McKenna et al. [2017b] integrate the entire visualization design process with visualization design worksheet, each sheet is featured with four generative methods and one evaluative method in sequence.

Workshops for data-driven storytelling include constructing and presenting narratives as a crucial part. Allen [2019] suggests a series of activities and approaches suitable for a workshop setting, e.g., simple narrative logic to be applied in stories such as “chronologically” or “leading with the ending”, and test the story with “three-minute storytelling”. Feigenbaum and Alamalhodaei [2020] compile the Data Storytelling Workbook with worksheets, mini-exercises, and activities to facilitate the communication of data independent from any specific process model to pursue a holistic understanding of data stories. For example, using “Mind the Gap” worksheet to identify indented message by the gap between audiences’ currently “Think, Feel, and Do" and the expected result; answering a list of “Five W” questions with paired activities to connect audience and valuable information; providing thinking directions to composite characters for data stories; addressing data bias in storytelling with standpoint approach; creating a 6-panel data comic in a 2-hour workshop; providing creators with an overview picture of necessary components in visual storytelling by “Four pillars” (i.e., symbol, color, caption, and editorial layout) and design principles. The activities introduced in this book focus on tasks for individual questions or certain stages of visual story creation. There are currently no structured activities for data-driven storytelling with data comics. Chapter 5 of this thesis describes a workshop plan for materials informed by the workshops and pedagogy above, the procedure is then summarised and presented as a structured model to inform the creative process, systems and tools for creating data comics.

2.5 Systems and Tools to Create Visual Data Stories

Systems and tools have been developed to support narrative visualizations from different aspects, and components, through manually adding or automatically generating. Researchers have investigated interfaces for non-programmers, supporting direct manipulation and clever data binding methods. This section reviews tools that facilitate visual data story creation, including adjusting and presenting multiple visualizations in multiple views, building transitions, and adding annotations.

Tableau [1] allows the export of interactive visualizations into a specific layout, resembling

[1] https://www.tableau.com/

comics or dashboards. Visual elements in visualizations are highlighted on mouseover and show tool-tips. Story Points in Tableau gives authors the ability to present a narrative by breaking a story into pieces based on different visualizations or different views of one visualization, in which the audience can read through a narrative by a navigator [Ryan, 2018]. Similarly, to break a story into multiple views, Lundblad and Jern [2013] integrate storytelling with geovisual analytic using HTML5, which displays dashboard associated storytelling panels.

Presenting a complete and coherent story requires the creator to assemble the story pieces (i.e., single facts) together with necessary transitions. Wohlfart [2006] presents a volumetric storytelling prototype application, the storyline is created by recording and editing the exploring process with interactive volume visualization rendering. The authoring system is developed by a storytelling model that contains a hierarchy of levels of story components including story node, story transitions, story action group, and story action atoms. For presenting smooth transitions, GraphScape [Kim et al., 2017d] focuses on automated sequencing by the similarity between visualizations and gives visualization design alternatives based on a given chart. Ellipsis [Satyanarayan and Heer, 2014] creates a stage element for each visualization, where creators can change visualization parameters, add annotations, and define scene transitions by setting conditions. Automatic generation with Calliope [Shi et al., 2021] can generate visual data stories and fact sheets from uploaded data.

Temporal sequencing is a common logic to tell through a series of events and facts. Tools are provided to create a temporal narrative in geo-spatial map [Eccles et al., 2008]; create an interactive timeline, navigating between slides with text, image, and videos [KnightLab, b]. There are also automatic solutions such as TimeLineCurator [Fulda et al., 2016] to generate a timeline with events through natural language processing [Cruz and Machado, 2011].

Annotations play an essential role in narrative visualizations. Creators can produce interactive annotated line graphs by uploading data sets in a spreadsheet template with Storyline Js [KnightLab, a], or use automatic solutions with Contextifier [Hullman et al., 2013a] for producing annotated line graphs of stock performance. Working with more generic visualization types, Aisch [2015] present Mr Chartmaker, supporting inline text-editing, directly selecting data points to add text annotation. Gao et al. [2014] develop NewsViews, which can extract annotation content by ranking the topical relevancy and annotating the descriptive sentence on the generated thematic maps.

2.5.1 Creating Comics and Data Comics

Crafting comics are currently supported through a range of diverse tools, some target comics or data comics directly, others are for more general graphic designs. Besides the mainstream graphic design tools such as Adobe Photoshop, Adobe Illustrator, or design tools like Figma, an array of tools particularly to create traditional comics exist. For example, Clip Studio Paint [Cli, 2013] provides a set of toolkits and features for comic drawing, including layout templates, environments, and characters. Comics can even be created without drawing, through tools that support creation by drag and drop of assets from a library [Com, 2021, Tobita and Shibasaki, 2009, Gramener, 2019], automatic layout solution [Cao et al., 2012], recommend ways to layout and generate a narrative of data story in comic strips [Zhao et al., 2021], presentation of exploratory analyses in a computational notebook as a comic [Kang et al., 2021], and creating comics from other media formats [Uchihashi et al., 1999, Tobita, 2010, Durrant et al., 2011, Chen et al., 2013, Zhao et al., 2015a]. To promote innovation and expressiveness with visual data stories, software and hardware are developed with hand-drawing sketching techniques by pen and touch interactions, e.g., DataToon [Kim et al., 2019], SketchStory [Lee et al., 2013], DataInk [Xia et al., 2018] and Geological Storytelling [Lidal et al., 2012]. To support more flexible workflows, Bigelow et al. [2017] provide a bridge between Illustrator and Data-Driven Documents [Bostock, 2019]. However, creating data comics with design tools requires significant manual work to make data visualizations or switching between graphic design and visualization creation tools (e.g., Tableau and Data illustrator [Liu et al., 2018]), and existing tools that target data comics are either only focused on specific types of data or provide limited freedom for expressive creation.

This chapter reviewed the background of visual data stories, comics for communication, learning, and teaching data visualizations and data-driven storytelling, creation methodology, and tools for creating visual data stories. The existing design methodology and tools for creating data stories and data visualizations provide abundant resources, that can inform the methods of creating data comics. While the past research suggests the potential benefits of using comics to explain science and statistics, little is known about the effectiveness of using comics for presenting data stories, and there is no sufficient design guidance for data comic creation. We believe that a better understanding of data comics can inform their creation and tool development, and providing effective creation supports will promote the production of data comics and bring benefits to more creators and audiences. By exploring the potential of data comics and their creation, the contribution of this thesis will shed light on how to use comics for data-driven storytelling.

Chapter 3

Towards a Data Comic Creation Model

As highlighted in chapter 2, although comics have been used in explaining science and for educational purposes, using comics for presenting visual data stories is still an emerging approach. To fill the knowledge gap of understanding and using data comics, the research of this PhD begins with understanding the effectiveness of using comics for narrative visualization, then moves to methods of using and creating data comics.

This thesis explores data comics from multiple perspectives. Inspired by comic as a narrative device, data comics involve elements in comics (e.g., *panel*, *transition*, *layout*, and *characters*), and elements in visual data narratives (e.g., data and visualizations). A data comic may or may not include all of the elements above, yet some are essential. This thesis further explore the *four essential elements* of data comics described by Bach et al. [2017a], including how the layout and the integration level of *words and pictures* affect comprehension and engagement (Chapter 4); using graphic depictions for explaining *visualization* techniques (Chapter 6); suggesting data comic as compelling solution for communicating *narrations* in scientific reports; demonstrating how *flow* varies with and without interactions (Chapter 8); and a plan of data comic creation in a workshop setting (Chapter 5). During the three years, the author and his collaborators explored data comics by learning from the literature of comics and visual narratives, creating data comics for different scenarios, evaluating creation methods with participants in workshops and collaborative design sessions, and devoting the effort to elevate the understanding of data comics to a new level.

Currently, there is no structured approach helping novices and professionals in design, illustration, or data science to create storytelling, especially, make tacit knowledge explicit, describe the relationship among the essential elements, and provide structured guidance for creation. This chapter introduces a *Data Comic Creation Model* to 1) describe the structured process of creating data comics, 2) elaborate on the theoretical setting and choices of the

methodology of this PhD, 3) map the contribution of each chapter to a big picture of creating data comics, and 4) evoke future research questions, tools, pedagogy for creating data comic and potentially other genres of data-driven storytelling with visualizations.

Towards answering the research questions raised in chapter 1, mixed research methods were used in this thesis, empirical evidence was collected and analyzed quantitatively and qualitatively. Quantitative data collected from a controlled study in a laboratory and wild setting provided evidence of effectiveness with the controlled conditions. Qualitative methods such as interviews and workshops were applied to explore why certain conditions make data comics engaging and the methodology of creating them. Data comics are man-made artifacts like other formats for storytelling. As few examples exist, the author of this thesis begins with creating data comics, then iterated, and evaluated the design with other participants. To evaluate the creation methods, and reach out to more creators and ideas, hands-on workshops and collaborative design are frequently used methods in this thesis. Those workshops also provide chances for the instructor to observe the participants when they practice, implying the encountered challenges, the methods they used, and how participants used the techniques and approaches we introduced.

Design models and the process is vast 2.4, methodologies for crafting data stories have been introduced by some researchers and practitioners. The Data Comic Creation Model is inspired by the five-plane framework of user experience design described by Garrett [2010] (Figure 3.1), the work-flow of making visual shared stories described by Lee et al. [2015] (Figure 1.5), also combined our experience of designing data comics, and the data comic creation workshops we conducted in the past. Instead of demonstrating an entire storytelling process from data analysis to presenting the story to audiences, the Data Comic Creation Model focus on the story making process (i.e., after having explored raw data and before delivering the final presentation), and particularly, with data comics. The Data Comic Creation Model described in this chapter was not published at the time this thesis was constructed, but the author of this thesis introduces it as a theoretical knowledge framework, and a reflection of what we have learned from comic creation and its facts. Having this model could foster systematic learning and teaching data-driven storytelling, designing workshop plans, development of tools, as well as informing future research in activities and evaluation.

3.1 The Five Layers of the Data Comic Creation Model

Currently, the Data Comic Creation Model consists 5 layers representing the key elements of creating data comics: I) Communication Objective, II) Message, III) Narrative structure and Explanatory Visualizations, IV) Visual Narrative Integration, and V) Graphic Appearance. The

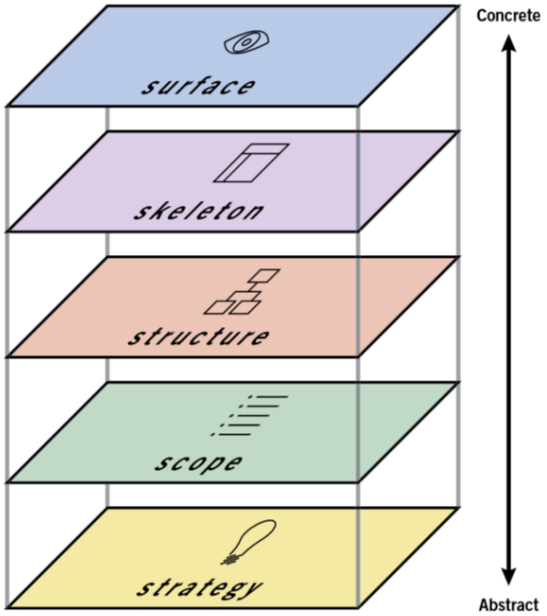


Figure 3.1: The Five Elements of User Experience Design by Garrett [2010]. There are five dependent layers, each level builds on the level before it, and they start with abstract level towards concrete level (from bottom to top).

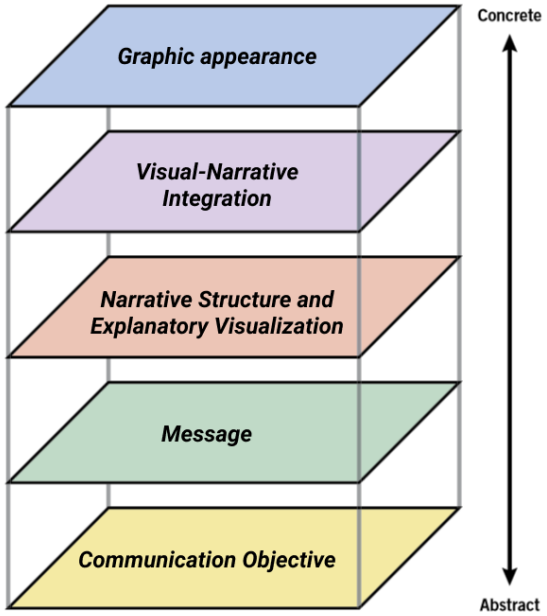


Figure 3.2: Data Comic Creation Model: A conceptual working model represents important components of data comics in the creation process. The five layers reflect five dependent stages of creating data comics as bottom-up sequence.

layers of this working model present the creation steps in a linear sequence, yet a creating process does not have to start from the bottom, and it can contain many loops, and tasks run parallel. The details of each layer are described as follow:

Communication Objective: The bottom layer indicates the communication objectives of the data story creator. This layer is informed by the "Strategy" of designing a product described by Garrett [2010], concerning both what product builders want to accomplish for their organization and what users want to get out of it. Likewise, an explicitly defined communication objective based on an understanding of the position of the storyteller, the data to be communicated, and the target audience can provide a reference for decision making in further steps and keep collaborators on the same page of understanding the work.

During this PhD, we worked with collaborators and participants from different domains to create data comics targeting various audiences and communication objectives. We found that creating data comics is conversations between the storytellers and their audiences with diverse knowledge backgrounds, levels of data, and visualization literacy. For example, the comic creators and their target audiences in the workshop (Chapter 5) are both data visualization

novices, the communication scenarios are mainly casual with light information load, for instance, as the creator of the comic in Figure 5.8 stated, “my audience are coffee buyers in coffee shops. My comic is available to read in coffee shops whilst you wait in line to get your morning brew."; data visualization cheat sheets (Chapter 6) support data visualization experts to explain complex visualization techniques to visualization novices “by providing a concise and consistent set of explanations for understanding specific visualization techniques, their visual marks, and visual patterns"; while data comics for reporting controlled user studies (Chapter 7) are meant to promote conversations between data and research professionals, or experts explaining their research to non-experts by “providing a framework for researchers and science communicators, guiding the dissemination process and designing visual narratives for scientific study reports". These chapters demonstrate how data comics vary by the conversation between different groups.

To deliver a successful user experience with a product, Garrett [2010] suggests “user segmentation” for understanding user’s needs by dividing the audience into smaller groups of people with shared characteristics. When delivering a data story, Lee et al. [2015] stress the audience, setting, and medium as “external factors” that influence how the story material will be created. To help with defining those factors, we suggested participants in our data comic creation workshops think about the following questions:

1) What is the story about, and under what context?

2) Who is the target audience and who is the decision-maker?

3) Where is the story likely to be delivered?

4) What does the story creator need their audience to know or do?

These four questions were asked foremost in the creation process. We found answering the four questions can help participants to understand the challenges, narrow down from exploration, and evaluate their works as criteria of self-assessment. In Chapter 5, we suggested participants to specify their target audience by *persona*, and specify *scenarios* (e.g., classroom, street, etc.) with scenario cards [Huron et al., 2017] before creating data stories. Other methods of understanding the target audience and communication objectives including “partitioning a large audience into smaller subsegments” [Duarte, 2010, p. 58], using activities like *Listening for* and *Mind The Gap* to identify the thoughts of target audience [Feigenbaum and Alamalhodaei, 2020]. Those methods are potentially helpful, yet have not been tested in our workshops due to time limitations.

Message: This layer indicates the understandable message supported by data. Towards creating a data story, the messages here reflect the derived data insights, and “data excerpts” from data exploration [Lee et al., 2015]. This layer aims to prepare a list of selected messages that provide detailed resources for *What does the story creator need their audience to know or*

do from the *Communication Objective* layer. In chapter 5, we split the message into *high-level messages*, *facts* and *data-literacy knowledge* for participants to specify the message needed for building a narrative in the next stage. The three types of messages reflect a thinking direction from abstract to concrete, and concern the needs and data literacy of target audiences. In our previous workshops, we encouraged the creators to write down a list of messages with bullet points, or on post-it notes, so that they could move them to find clusters and relations among individual messages, and discard the ones with less relevance. While not all of the messages from the list finally go in the story, especially the high-level messages, that are expected to be taken as conclusions with the audience. For instance, in Figure 5.8, no text such as "Attention is needed on species diversity" nor "species extinction is a growing problem" are found in the comic, yet an audience could get this conclusion by data of one type of frog and a character who decides to escape from the fate of extinction.

Including or excluding messages is a decision of trade-off. It is challenging to select messages and facts from a sea of relevant information. Including too many messages takes extra space and makes the comic visually overwhelming, while too little information leaves a story less informing, or not sufficient enough to support the high-level message the creator wants to convey (Chapter 4). To inspire creators with messages to be included when reporting a controlled user study with data comics, we provide a structured framework with a list of messages in each stage in Chapter 7 (Section 7.4). We found the message selection was still a hard decision, determined by multiple factors such as requirements from stakeholders, interests of audiences, and restrictions of space and time of presenting, referring to the four questions from *Communication Objective* layer.

Narrative Structure and Explanatory Visualization: With a list of messages prepared from the previous stage, the next step is to assemble the messages and develop a compelling narrative accompanied by explanatory visualizations. This layer has the same goal and tasks as the "Make a Story" in the visual data storytelling process, in which creators construct the storyline or plot. A narrative structure, for example, Freytag's Pyramid [Freytag, 1895], Martini Glass Structure, Interactive Slideshow, and Drill-Down Story [Segel and Heer, 2010] are commonly used in data stories to impress the audience and affect their perception and understanding of data. The narrative structure is also studied with comics, with contents of panels categorized by *Establisher, Initial, Prolongation, Peak* and *Release* [Cohn, 2013b]. To help participants to build a narrative, we suggested they tell a "3-minute story" in their groups inspired by Allen [2019] (Chapter 5). While in many cases, stories are not often told linearly but use flashback and other patterns [Riche et al., 2018], we introduced interactions to static data comics in Chapter 8, which enable more possibilities with narrative structure.

Besides a narrative, data visualizations provide data-based evidence that supplements the messages of the story. Data visualizations offer data-based evidence for tendency, relations, outliers, clusters, and distributions. The visualizations used in visual data stories are focused on communicating messages and promoting understanding. To help creators with explaining data visualizations, we present data visualization cheat sheets in Chapter 6, we hope the graphics-based explaining approach could inspire creators when interpreting visualizations in data comics or other visual data stories.

Visual-Narrative Integration: This layer aims to integrate the narrative and visualizations prepared from the previous stage and break the story into smaller pieces with necessary visual supplementation, so that makes the story easier to consume and the important messages prominent. According to our findings, there are several factors to be considered:

- The *granularity* of information pieces, ranging from one message per panel to multiple messages per panel. For example, the comics used in the lab study in Chapter 4 consist of one message per panel, while comics presented in Chapter 7 contain more than one message per panel to reduce space and repetition. While the effective difference of low and high information granularity in data comics requires further study.

- The *pace and rhythm* the creator wishes to present. The number of panels and pace can affect the rhythm of storytelling. For example, to gradually reveal a chart by splitting it into several sections in a few panels or to show the entire chart at once.

- The *transitions* between panels influence the flow of the story. Consistency between states is likely helpful for viewers to generate the explanations that tie the patterns represented by visualizations into a coherent story [Hullman et al., 2013b]. Data comic design patterns suggest transitions for different narrative purposes such as Temporal changes, Zoom, and Overview & Details [Bach et al., 2018c].

- The *layout* of a comic concerns how panels to be positioned and grouped, which can affect audiences' reading sequence and their understanding of the association between messages. Comments from our study participants confirmed the advantage of using comic panels and layout to propel the narrative and group information by chunks, rows, and pages (Chapter 4 and 7).

- The *Narrative guide* is another tool that is commonly used in comics to guide the audience through the story or to express a point of view on the events recounted to distinguish it from the real creator [Groensteen, 2013]. Yet the use and influence of characters in data comics require further research guidance in future work.

Graphic Appearance: At the top layer of the Data Comic Creation Model, the *Graphic Appearance* focuses on the visual arrangement and appearance such as the visual salient [Itti and Koch, 2001], styles, font, and color usage. Many literature attempt to give graphic design guidelines for data visualization (e.g. [Koponen and Hildén, 2019, Iliinsky and Steele, 2011]), and explore various styles of comics (e.g., [Madden, 2005, Molotiu, 2009]). Currently, there is no guidance for designing these elements in *Graphic Appearance*, the visual design of the data comics presented in this thesis are mainly based on the creators' experience and their understanding of the story, *Communication Objective*, and personal preference. We find the design space of *Graphic Appearance* is challenging to define, similar to styles in paintings and data visualizations [Koponen and Hildén, 2019], *Graphic Appearance* is not only about conveying the information, but also emotion, feelings, and personal style of the creator.

3.2 Limitation

The Data Comic Creation Model is developed to streamline the creation of data stories and complement existing models. Similar to many other high-level process models, the workflow presented in this model provides one yet not the only way of thinking and creating data comics, which means it can not generalize the creation process of data comics, the best practice still requires situated acting. We acknowledge that the stage of building a data comic is often emerged with looking for and exploring new data, the creation process can be more dynamic in the practice that the steps are not always separated clearly. Currently, the model is still at its initial stage, whether this model can lead to better results requires further evaluation.

Chapter 4

Comparing Effectiveness and Engagement of Data Comics and Infographics

4.1 Motivation

The effectiveness of communication is a basic criterion to evaluate a format for data-driven storytelling. However, to the best of our knowledge, there is no study investigating the effectiveness of communication with data comics. This thesis begins with a study comparing data comics with infographics. Generally, the use of comics in classroom education and health-related communication appears to improve the reader's comprehension and engagement [Negrete and Lartigue, 2004, Farinella, 2018]. While these results may encourage the use of data comics, generalizing existing study results to data comics is challenging due to varying content, visual representations, audiences, and purposes. In particular, graphics in data comics focus more on abstract data visualization [Bach et al., 2017a], require visualization knowledge, and largely deal with concepts requiring data literacy such as relations, temporal change, and quantities. For example, Qu and Hullman [2018] points to the importance of consistency in creating sequences of visualizations, a factor that very likely influences the acceptability and usability of data comics. Given the popularity of infographics and traditional visualization formats, as well as the body of expertise built up around their creation and use, it would be wrong to simply assume there will be enhanced understanding and engagement with data comics.

As a step towards a better understanding, we provide the first structured investigation into the effectiveness of data comics. We compare data comics with the two closest popular alternatives: infographics and texts accompanied by visualizations (illustrated texts). While there are numerous differences between these three formats, we focus our comparison on *i)* the degree to which an explicit *reading order* is given, and *ii)* how close the *integration of text*

and picture is (Section 4.2). While data comics typically have a highly structured reading order and text closely integrated pictures, illustrated texts and infographics exhibit a varying degree in each of these dimensions. We compared these techniques in a controlled lab study, where 36 participants read the same 3 stories, each presented in a different format (Section 4.3). We collected empirical and subjective data on understandability, recall, preferences, and engagement. To complement this controlled study, we observed 50 groups of visitors engaging with comics and infographics in an open public space (Section 4.5, Figure 4.8).

This chapter aims to support the further investigation, application, and creation of data comics through empirical results on their effectiveness. As far as we are aware, this is the first study to address this gap. Our results are encouraging for the use of data comics, especially for complex spatio-temporal data, which are naturally hard to visualize in infographics. Our findings bring knowledge to the layer of *Visual-Narratives Integration*, also lead to valuable design guidance and design consideration of the trade-offs.

The original publication of this chapter collaborates with Shunming Wang, Matteo Farinella, Dave Murray-Rust, Nathalie Henry Riche, and Benjamin Bach [Wang et al., 2019c]. The author of this thesis prepared the materials used in the lab study, including modifying and designing the three formats (data comics, infographics, and illustrated texts); took the responsibility of running the study; took part in analyzing the data, and writing the paper.

Infographics and Illustrated Texts

Two popular alternatives to data comics are *infographics* and *illustrated texts* ("magazine style" [Segel and Heer, 2010]). Both are text-picture combinations, differing in how these elements are combined. Infographics strongly emphasize the visual content, often using visual embellishments to stimulate attractiveness and memory [Bateman et al., 2010]. They tend to be open-ended, inviting the reader to explore the content without any specific direction. For more structured narrative content, e.g., in journalism and scientific papers, texts illustrated with visualizations have become the common option: a text for the main narration, referring to visualizations as needed.

All three formats (comics, infographics, and illustrated texts) are expressions of different combinations of reading order and text-picture integration. Current studies are not conclusive and say little about these three formats' ability to support understanding and memory as well as how engaging each format appears to potential readers.

4.2 Study Design Space

Infographics, illustrated texts, and data comics vary in a range of characteristics and it is not easy to precisely define each format. Thus, to inform our baseline techniques for the study, we focus on two characteristics across these formats:

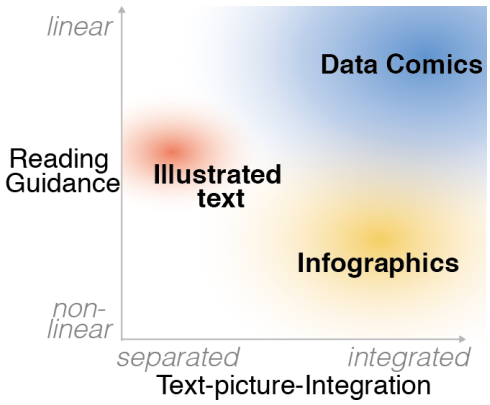


Figure 4.1: Study design space

1 Text-Picture-Integration is the *spatial distance between the verbal message and the visualization* [Bach et al., 2017a]. In their general form, comics exhibit a high integration of text and picture; one piece of text is related exactly to one image, both supporting the same information within the panel. A low integration would mean that text and picture are presented in parallel with occasional implicit or explicit references such as found in scientific and news articles.

2 Reading Guidance is *how strong the reader is guided while reading the story* [Bach et al., 2017a, 2018c]. While exceptions exist [Bach et al., 2018c], comics provide high guidance through an explicit order of panels. Low guidance can be found in most infographics where text and pictures are usually not organized in a linear order but can be read in a non-linear way.

We use these dimensions to map out a design space for visual storytelling formats for our study (Figure 4.1). *Comics* are located high on both dimensions, with tight text-picture integration and strong reader guidance. *Infographics* provide a less strong integration of textual and pictorial content because the same picture is usually annotated with several textual information and the specific scope of the text, with respect to the picture. Infographics provide low reading guidance as the content can usually be read in any order. *Illustrated Texts* are by visualizations, sometimes referenced in the text. This format provides a low integration of text and picture, a higher reading guidance than infographics, but less than comic as it is up to the reader to switch between text and picture.

To investigate the effectiveness and engagement of these formats, we opted to perform two complementary studies: one in laboratory settings and one in the wild.

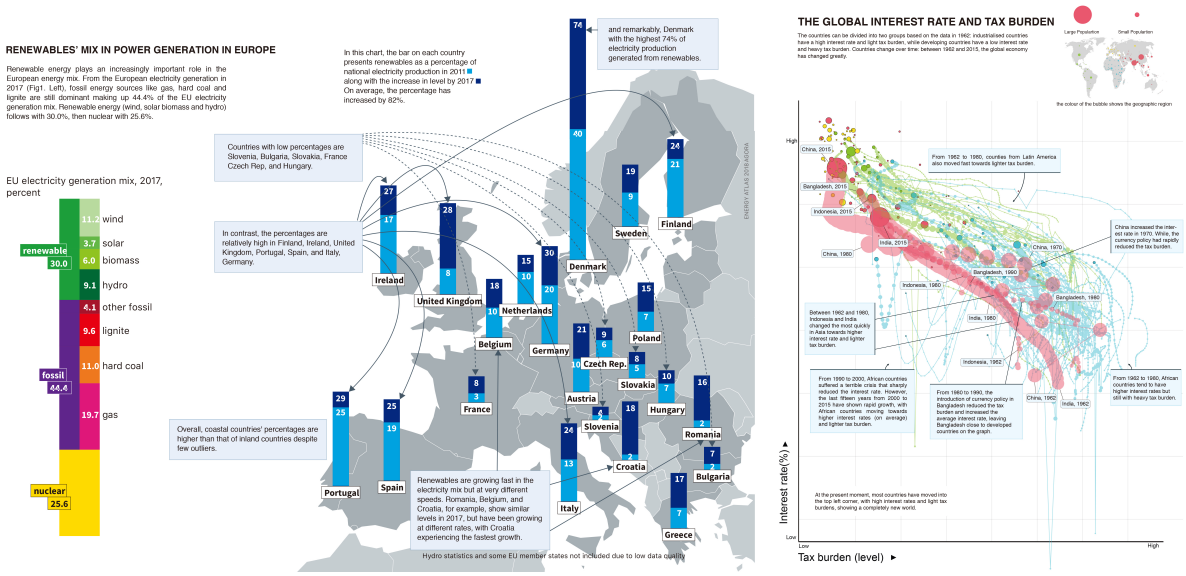


Figure 4.2: Infographics: ENERGY (left) and ECONOMICS (right).

4.3 Lab Study

Our first study aimed to compare comics, infographics, and illustrated text in a controlled lab study. We were generally interested in how each of the dimensions, laid out in the previous section, influences comprehension, recall, and engagement with the formats. More specific questions concerned how people read each format, which configurations they found helpful or confusing, how well did they understand the communicated stories, as well as how much of the content do participants remember after a while. To that end, we invited participants to read a series of stories in each of the formats, answer questions related to content understanding, and report their subjective impressions about each format. Prior to the study, we carried out a pilot with three visualization-experienced participants to ensure the suitability of our material and techniques. The full material (including all questions and results) is found in the supplementary material.

4.3.1 Stories

During the study, participants read three data stories, a number we chose with respect to how long we expected participants to remain attentive and remember story content. Our main criteria in finding appropriate stories for a controlled lab study were (i) comfortably presentable on A3 paper and (ii) requiring little previous knowledge. To extend the variation among the study examples, we selected different types of common visualizations including multivariate data, dynamic networks, and geological distribution. To ensure realistic stories, we selected three

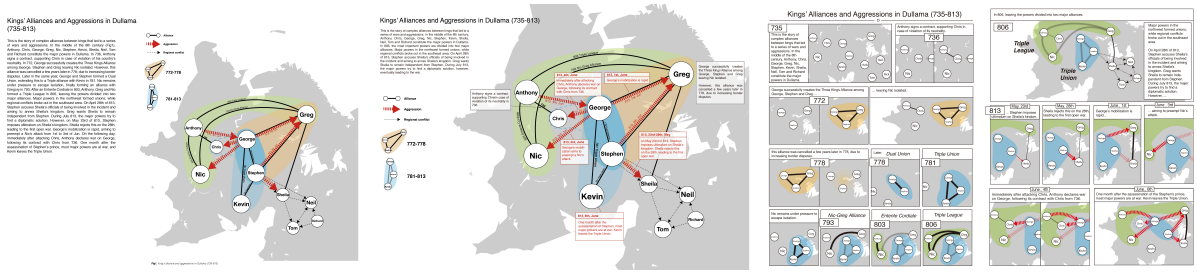


Figure 4.3: Samples of ALLIANCES in the lab study. ILLUSTRATEDTEXT (left) ; INFOGRAPHIC (center); COMIC (right).

existing infographics (see below). For two of the stories, the labels, text, and data were changed to prevent users from relying on prior knowledge.

- **ENERGY:** ***Renewables' Mix in Power Generation in Europe*** talks about the production and distribution of renewable energy among European countries. The visualization was found in the Energy Atlas [Bertram and Primova, 2018]. Bars on each country showed production in 2011 and 2017 respectively. Texts describe features of the distribution (Figure 4.2-left).

- **ALLIANCES:** ***Kings' Alliances and Aggressions in Dullama*** (Figure 4.3) was based on a graph comic on dynamic networks *European Alliances before World War I* [Bach et al., 2016a]. Dates were changed and the story was transposed into a fictitious medieval kingdom. Country names were replaced by common names and the background map was altered.

- **ECONOMICS:*Global Interest Rate And Tax Burden*** was taken from Hans Rosling's TED talk [Rosling, b], featuring a dynamic scatterplot of country indicators over time [Rosling, a] (Figure 4.2-right). We changed axes labels to *Interest rate* and *tax burden* respectively, implying a fuzzy meaning. The narration was guided by the changes to specific countries over time.

While we tried to keep stories to an equal level of complexity (e.g., number of panels, number of messages per panel, expected commonness of visualizations and difficulty of understanding visualizations), all three participants in the pilot study gave the same ordering in terms of increasing complexity: ENERGY, ECONOMICS, and ALLIANCES.

4.3.2 Techniques

The techniques in the lab study are the three techniques mentioned in Section 3. As boundaries between the formats are blurred, we settled on the following criteria to create the material for

our comparison. Figure 4.3 shows the three formats for ALLIANCES. The full material for the other stories is available in https://datacomics925658343.wordpress.com/study/.

- **ILLUSTRATEDTEXT**: One or two narrative paragraphs were placed right below the title, followed by the legend below or on the right. Text and pictures were visually separated, captions and essential names were added on the pictures referring to the text description.
- **INFOGRAPHIC**: Besides a 1-2 sentence summarization under the title, texts were shown in text boxes, placed close to the visual information they referred to. Leader lines sometimes connected text to visual elements. No explicit visual hint implied a specific reading sequence.
- **COMIC**: All text and visuals were placed into 16-18 panels in an unambiguous linear left-right, top-down order. Visuals were adapted to support the text, e.g., focusing on specific information, highlighting important and removing unimportant information. We made no other changes to visual encoding or text.

We ensured everything was readable comfortably from 40cm distance, printed on A3 paper format. Texts were kept the same across formats, except minor adjustments to the format. For example, we added *"see figure x"* for ILLUSTRATEDTEXT and simplified few longer sentences to fit the panels in COMIC. We created all materials through several iterations to make sure they were readable and understandable and to lower confusion; at each stage, we discussed in depth across the authors and asked for feedback from three external comic artists as well as two external colleagues. The pilot study identified minor problems in understanding and design and we consequently modified several ambiguous expressions in the stories and the questionnaires. To provide a high-resolution and minimize the operational distractions such as panning and zooming, we printed each story on one-side A3 paper instead of showing them on a monitor.

4.3.3 Hypotheses

We developed our hypotheses based on the existing literature, and the two dimensions of our design space.

- **H-ACCURACY**: We expect participants *to have a more correct understanding of the messages* after reading COMIC than INFOGRAPHIC, and least with ILLUSTRATEDTEXT. We believe that making a connection between visual and verbal for each message improves understanding. Comics with messages embedded in individual panels will be more effective than combined information in INFOGRAPHIC and separated text and pictures.

- **H-ENGAGING**: we expect COMIC *and* INFOGRAPHIC *to be rated as more engaging than* ILLUSTRATEDTEXT. This means that people are more willing to spend time with COMIC and INFOGRAPHIC and believe they are more fun to read.

- **H-MEMORABILITY**: We expect COMIC to increase retention of information compared to the other formats due to the separation of information into clear chunks, supported through tightly integrated texts-and-pictures, presented in a clear reading order [Mayer and Gallini, 1990].

4.3.4 Data Collection

Seven types of data were collected during the lab study: (1) error rate from multiple-choice questions to measure understanding of each story and format (including four answer possibilities and "I'm not sure"); (2) story recall scores, coded by the experimenter and explained in Section 5; (3) reading paths drawn onto the material by the participants; (4) subjective scores, and (5) qualitative feedback from participants on all three formats for the same story (shown at the end of the study) gathered through semi-structured interviews. Questions included *"Compared to the other formats of same story, how would you judge these formats?"* and *"Do you have any preference among these formats and why?"*. Subjective ratings were collected through a questionnaire after participants saw the three formats of all stories. While engagement is hard to define and measure [Mahyar et al., 2015], we assessed aesthetic feeling, willingness of spending time to explore, enjoyment, attention, engagement and preference using 7-point Likert scales, inspired by existing frameworks [O'Brien and Toms, 2010, Saket et al., 2016].

4.3.5 Participants

We recruited 38 participants through mailing lists of a Western-European university. One subject was eventually denied participation due to missing English language proficiency and one subjects results were invalidated as she did not complete phase 1. Of the remaining 36 participants, 19 were male, ages ranged between 18 and 35 years, 1 A-level, 9 undergraduates and 20 graduates, 4 computer science doctoral students, 1 research associate in computer science and 1 lecturer of Linguistics and English Language. Participants' backgrounds included Design, Art, Computer Science, Engineering, Linguistics, Philosophy and Psychology. We had 16 Europeans, 15 Asians, 3 North American and 2 African; 34 had lived in an English speaking country for more than 1 year, the other 2 had scores in English language test equal to B2 level of Common European Framework of Reference for Languages.

To assess participants' level of reading ability, comic and visualization literacy, self rated questions were filled in before the study. For text reading ability, 13 read illustrated texts daily and 13 indicated weekly; 11 read scientific text-only articles daily and 12 read them weekly. For infographics, 20 read them weekly and 7 daily. Familiarity with data visualizations was rated differently (**1** **7**). For comic literacy, 8 read comics daily, 7 weekly, 4 monthly, 9 few times year, 3 yearly and 5 of them have never read comics with 20 of them having first read comics aged between 6 and 12 years old. 12 participants had experience in drawing comics, but mostly only a few times and for less than one year. Participants reported on their preferred learning method: verbal (spoken or written) media (6), visual stimuli (8), combined visual and verbal media (17). Repetition was a popular strategy mentioned by 8 participants. Participants were paid a compensation of $7.

4.3.6 Procedure

Our study employed a mixed design. For both stories and techniques we employed a within-design, i.e., every participant read all stories and all formats. However, as we could not provide the same story twice to the same participant, we employed an in-between design for story-format conditions, i.e., every participant read each story using a different format. We used Latin-square randomization to assign stories to formats. With 36 participants each seeing 3 presentations, each of the 9 story-format combinations (3 formats $\times$ 3 stories) was seen by 12 participants. The experiment was divided into three stages, taking place on different days. The first part (lasting between 45 and 55min) included the following steps: (1) Upon arrival, participants filled out a background questionnaire asking for, e.g., familiarity with comics and visualization. Participants sat in a quiet room with the instructor. Material was printed and placed on the table, one at a time. (2) Participants read each story in a fixed order (ENERGY, ALLIANCES, ECONOMICS), each time with a different format. Participants were told to read comfortably without rush but to try to not spend more than 5min reading, and that they were asked questions about their understanding after reading. (3) Immediately after reading a story, participants were asked to re-tell the story with their own words. Reading and retelling was recorded through a video camera, not including the participant's face. Then (4), they were given a questionnaire with 6 multiple choice questions to assess participants' understanding (see next subsection). Eventually, (5) participants were asked to explain their reading strategy for each story in the format they had viewed, by drawing arrows onto the printed material. Participants came back the next day for their 2*nd* session, lasting between 20 and 30min each. For each story, participants were asked to (6) retell the story with their own words to assess recall. Then, (7) they were presented with all the materials

(COMIC, ILLUSTRATEDTEXT, INFOGRAPHIC) for each story—including the material they had not seen during their reading session—and were asked for their preferences through another questionnaire and then (8) were asked additional feedback on all three techniques in a brief semi-structured interview. Eventually, (9) 1 month after these sessions, we emailed participants the multiple-choice questionnaire form (4) to further help assess recall. Participation was rewarded with the possibility of winning a $25 Amazon voucher.

4.3.7 Understanding Questions

In step (4) participants answered 6 questions, measuring their understanding of the story content. Each question provided 4 answer options plus an *"I'm not sure"* option. Questions were naturally different for each story, yet for each story we created one question covering the following information types: **distribution** (e.g., *"What is the geographical distribution characteristics of renewables production in EU?"*), **time** (e.g., *"During what time did African countries suffer from crisis"*), **single fact** (e.g., *"The tax burden reduced in which country due to family planning according to this story?"* **outliers** (e.g., *"Which King was isolated when the Three King Alliance was created in 772?"*), **comparison** (e.g., *"Compared to 1990 to 2000, African countries tend to have (?) interest rate and (?) tax burden from 2000 to 2015." higher / lower*), and **visual encoding** (e.g., *"Which colour is used to present countries in Latin America"*).

4.4 Lab Study Results

All 36 participants completed the study, with each session lasting just below 1h. For each of the 9 story-format combinations, we obtained true or false answers to 6 questions from 12 participants. We collected all video materials and annotated print outs from each participants, available in https://datacomics925658343.wordpress.com/study.

4.4.1 Understanding

Understanding was measured as accuracy for the questions from the questionnaire in step (4). Error per question was binary, i.e., participants selected the right or a wrong answer. For each participant we calculated the mean accuracy score across all the answers per story. Using D'AGOSTINO'S K-SQUARED TEST, we found accuracy scores for two stories to be *not* normally distributed. Below and in Figure 4.4 we report on mean values, confidence intervals (CIs), and effect sizes. Using Bonferroni correction for multiple comparisons, we report on 98% CIs for

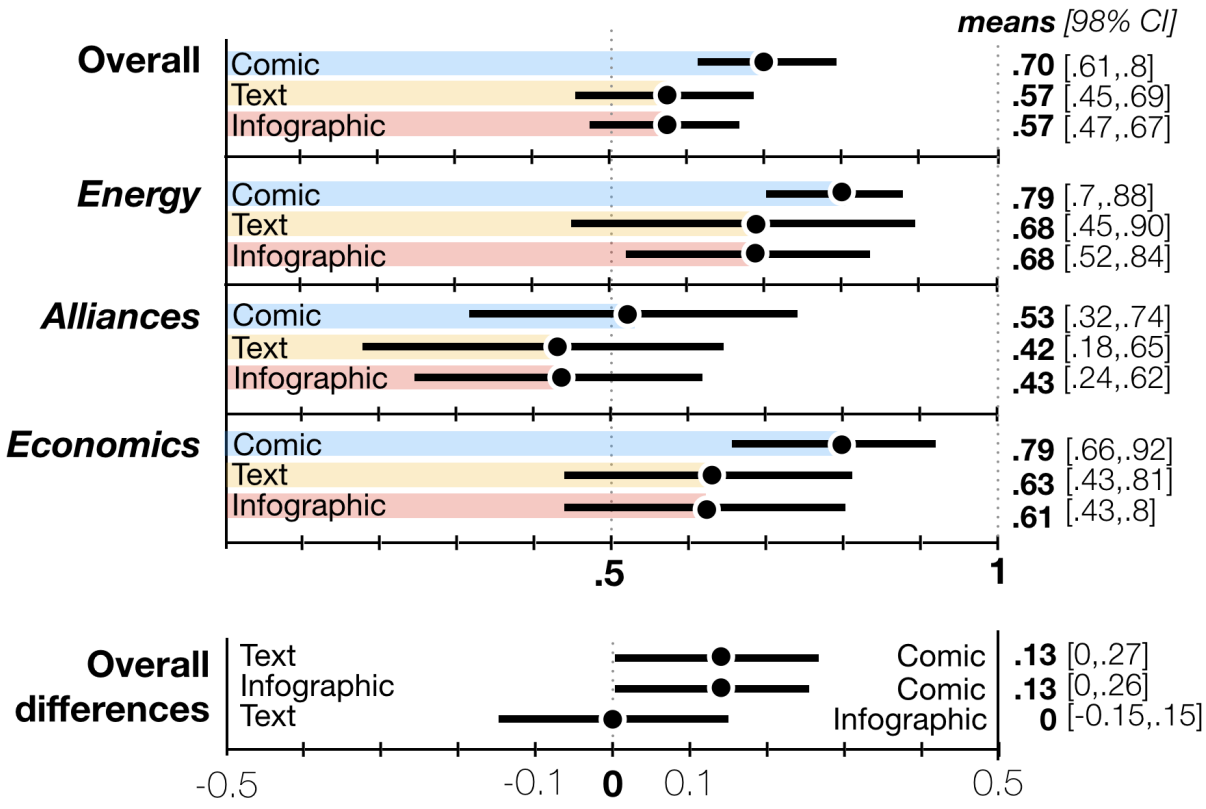


Figure 4.4: Results for *(top)* understanding including means and 98% CIs and *(bottom)*: overall effect sizes with 98% CIs.

our three comparisons ($1-0.05/3$). P-values are indicated for pair-wise comparisons yielding significance at the respective 0.02 level or close, using a MANN-WHITNEY-U test.

Across Stories we found COMIC (mean=.70) to be more accurate than INFOGRAPHIC and ILLUSTRATEDTEXT ($p<.006$). No difference was found between ILLUSTRATEDTEXT and INFOGRAPHICwith the same mean accuracy of .57. Effect sizes between techniques are reported in Figure 4.4-bottom. Our effect sizes represent the overall improvement (or decrease) in understanding averaged for each participant, i.e. for each participant we calculated differences between formats and averaged these values. Differences between COMIC and the other techniques amount to .13 points in understanding, implying that comics are on average 23% more accurate than both ILLUSTRATEDTEXT and INFOGRAPHIC.

For **ENERGY** we found COMIC (mean=.79) more accurate than INFOGRAPHIC (mean=.68). However, leading to the same mean accuracy than INFOGRAPHIC, ILLUSTRATEDTEXT was not found different from the other two techniques and showed much wider CIs. For one question, we could find a real difference between techniques; the question asked for spatial distribution of countries. As this information was highlighted explicitly in one panel in the comic, we believe it

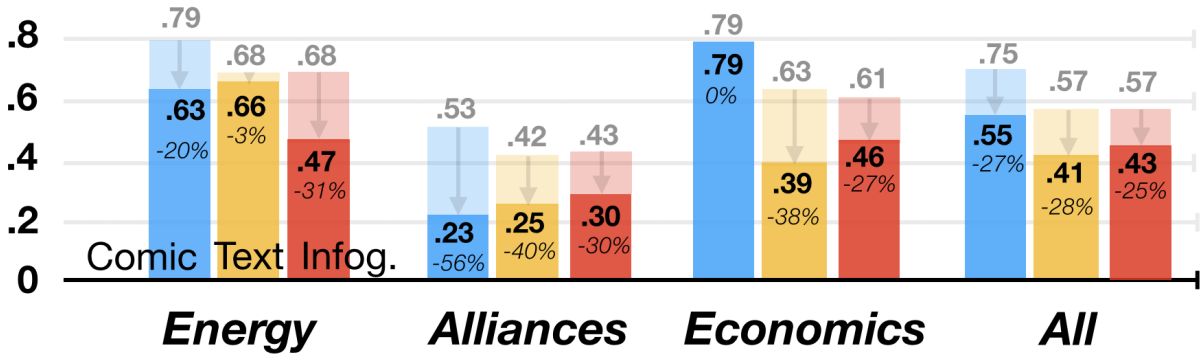


Figure 4.5: Mean results for understanding after one 1-4 weeks. Upper numbers indicate original values (as shown in Figure 4.4), lower numbers indicate correct results after 1-4 weeks with change in percent (COMIC=blue, ILLUSTRATEDTEXT=yellow, INFOGRAPHIC=red).

was easier for participants to understand and remember.

For **ALLIANCES** we could not find any clear differences, while COMIC (mean=.52) was still the most accurate on average but with largely overlapping CIs. INFOGRAPHIC and ILLUSTRATEDTEXT had similar mean values (.42 and .43). Again, we found a difference for one question asking about the meaning of a specific visual encoding in the visualization (*"What does the black dashed relation represent?"*); COMIC was significantly worse (mean=.25) than the other two formats (ILLUSTRATEDTEXT=.66, INFOGRAPHIC=.83). We believe participants overlooked this information in the comic as they might have paid less attention to this particular—not explicitly highlighted—detail. We attribute the good performance of INFOGRAPHIC to the fact that a respective text was placed close to these lines to explain their meaning.

For **ECONOMICS** we found similar results to ENERGY; COMIC showed a slightly higher accuracy (mean=.79) while INFOGRAPHIC(mean=.61)($p<.036$) and ILLUSTRATEDTEXT (mean=.63) ($p<.04$) were similar. As with the other stories, we found a difference for a question on time (*"What is the characteristic of the countries in 1962?"*). Here, COMIC was more accurate (mean=.91)($p<.015$) then any of the other formats (ILLUSTRATEDTEXT=.58, INFOGRAPHIC=.33). We conjecture that the bad performance of INFOGRAPHIC can be related, again, to the missing temporal visualization, while COMIC showed that information in a single panel. While the explanation should indicate a similar bad performance for ILLUSTRATEDTEXT, the average here was slightly higher. However, this particular infographic has been described as very cluttered, which may explain some of the poor performance.

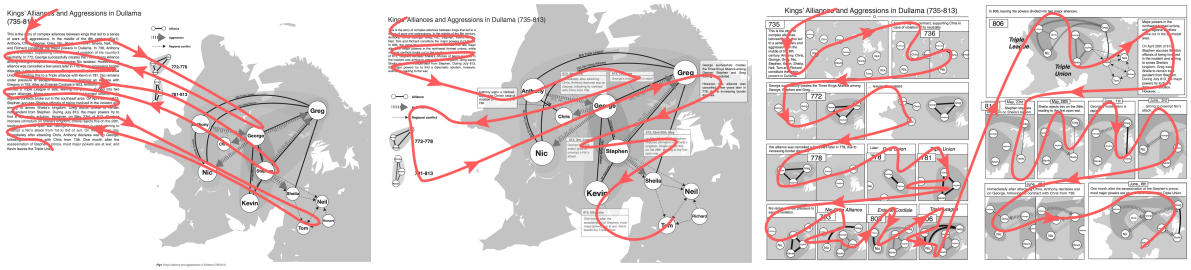


Figure 4.6: Examples of users' traces while reading the same story (ALLIANCES) in different formats. Percentages in brackets indicate the fraction of users showing comparable traces. Left: ILLUSTRATEDTEXT (67%), center: INFOGRAPHIC (75%), right: COMIC (100%).

4.4.2 Recall

After reading the stories on the first day, participants were required to retell the story in their own words. We took notes of each explicitly presented fact they mentioned (11 for ENERGY, 14 for ECONOMICS, and 17 for ALLIANCES). Self discovered messages, i.e., those not being mentioned in the texts, were not considered. On the second day, participants were asked to again retell the story with their own words. Following the methodology by Bateman et al. [2010], we coded the difference between both versions as follows: 3 points for every correct fact (e.g., correct values), 2 points for remembering general trends (e.g., increase, type of temporal change), 1 point for vaguely remembering (e.g., mentioning type of information) and 0 points for not or wrongly recalling. Two of the authors independently coded all of the recordings, then discussed until reaching consensus. We found COMIC to yield slightly more precise results on average (35%), compared to ILLUSTRATEDTEXT (32%) and INFOGRAPHIC (30%).

After one to four weeks (varied by the time each participants did the questionnaire for the first time), we sent the questionnaire from step (4) again to all participants (Figure 4.5): Overall, COMIC caused participants to remember most on average (55%, down from 75%), followed by INFOGRAPHIC (43%, down from 57%) and ILLUSTRATEDTEXT (41%, also down from 57%). Participants lost around 1/3 of their performance. However, values varied across stories, with each format performing best for one story. In two cases, mean understanding rates did remain the same (ILLUSTRATEDTEXT for ENERGY and COMIC for ECONOMICS).

4.4.3 Reading Strategy

By asking for the reading sequence of the story, we simulated an "eye tracking" phase (step (9)) in our study. In the pilot study, we found participants walking us through their reading order more accurate and informative for our purpose than actual eye-tracking. Figure 4.6 shows

example traces, drawn by participants, while explaining their reading order. With COMIC, all participants followed the panel order, reading both text and looking at figures in each panel except 1 participant who read the COMIC of ENERGY in a right-left order from the second row. 16 participants (44.4%) jumped back to previous panels to make comparisons. For INFOGRAPHIC, most participants (83%) started with title and abstract, the others started with figures. We found no specific sequence in which participants read text-boxes in INFOGRAPHIC, i.e. no prevalence for left-right or top-down order. We found reading sequence was guided by the layout of text boxes, e.g., clock-wise or randomly in ECONOMICS and ALLIANCES. Generally, participants were guided by text boxes and only then looked at the figure to obtain more information. All participants checked the legend in ALLIANCES and ECONOMICS. In ILLUSTRATEDTEXT, most participants (83%) started with the text, few (17%) started with pictures. 28% checked figures in the end, while a majority (72%) checked the figures during reading (2-3 times), especially when they found specific values in the text.

4.4.4 Subjective Feedback

Subjective results from the preference questionnaire (step 7) are summarized in Figure 4.7, following the same conventions and analysis as Figure 4.4. We found COMIC to be highest rated (averages) on three measures (*fun*, *engaging*, and allowing people to stay *focused*) while scoring slightly less then INFOGRAPHIC for *aesthetics* and *exploration*. While, differences between COMIC and INFOGRAPHIC are minor and not significant, COMIC is generally ranked higher as shown in the respective distribution (small bar charts left side of Figure 4.7). ILLUSTRATEDTEXT was rated generally least across all measures with huge differences to the two other techniques for the measures *engaging* and *fun*. INFOGRAPHIC was rated as *aesthetically pleasing*, with opinions being more equally distributed for COMIC. Asked which format participants would chose next time, we found similar results: COMIC rated best by 47% (**worst ▪▪▪ best**), ILLUSTRATEDTEXT rated least by 64% (**worst ▪▫▫ best**) and INFOGRAPHIC in the middle (**worst ▫▪▫ best**). In the following, we report on participants' feedback during the interview in step (8). Frequencies of issues raised are reported by the numbers in brackets. Full material and an overview table with the reported merits and drawbacks of each technique can be found in https://datacomics925658343.wordpress.com/study/.

COMIC—Comics were appreciated for their clear reading order (33%) and their ability to break down the complexity into pieces (28%). Reading order was found use support memory (6%)(*"[...] because the important information is repeated all the time, which helps [...] memorize."*,*"Comic makes a story in your head"*) and were found to facilitate understanding: *"just*

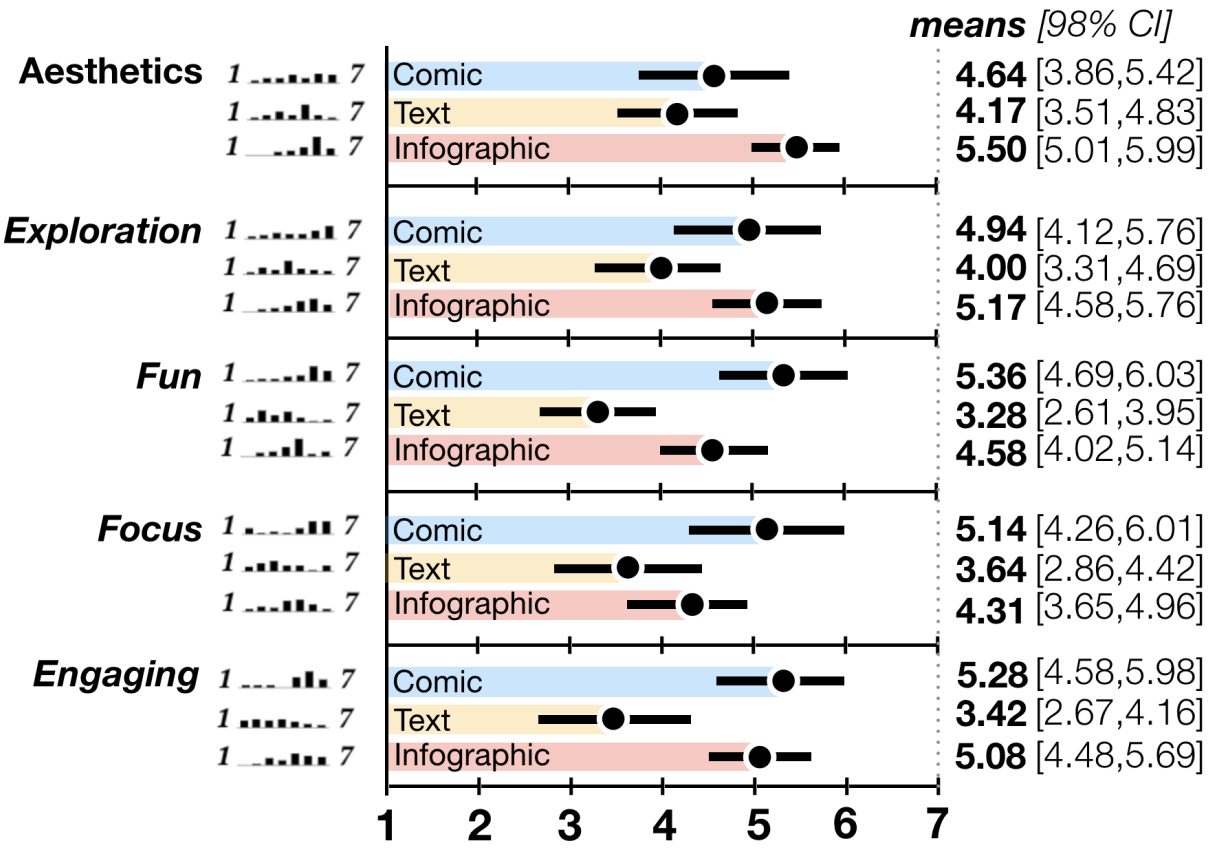


Figure 4.7: Subjective measurement from the lab study in step (2) of the 2nd session including means and 98% CIs. Distributions of the answers are shown on the left.

follow the sequence. It is logic and well organized", especially for temporal content (39%) (*"just because [*ALLIANCES*] is so complicated and chronological"*). Breaking down complexity into individual pieces has been found useful for the same reasons (*"You have the option to see the information by steps, you can easily have your memory when it is happened (sic). It is like the same way we remember history when I was a child." "is quite easier to memorize"*. Additionally, participants liked that comics could group higher-level messages into rows and potentially pages. Participants also commented on the ability to quickly overview information and find/recall the information they wanted (*"If I don't have time, I'll go for data comic"*). In fact, during recall, 4 participants used their hands to air-point to where the respective panel was located.

On the downside of COMIC, visual repetitions were distracting for 2 participants, indicating too much information to process: *"every time I see new pictures (panels) I expect something new"*,*"I need to compare to define what is new"*. There is tension here, as building a new message is on the basis of previous messages, yet to present previous messages by repeating the visual elements in panels can cause redundancy. For ENERGY, one participant mentioned *"I think it is unnecessary to have that kind of level to break down, because it is simple enough to*

understand the image by just looking at the whole picture and explore it.". This story indeed featured a high degree of visual repetition (the map of Europe) as it was hard to break down the content into simpler images. Again, other people explicitly preferred a more open format (*"I can only read the comic step by step, it is hard to find the part that I am interested in"*) and found that an overview picture was missing. The same participant was found to jump between panels. Looking at the comic the first time, one participant noted *"This one can be confusing, considering there are all types of graph. It seems a little much at first glance"*.

INFOGRAPHIC—Infographics were rated highly for exploration (33%) by providing overview and detail at the same time which helped to make comparisons (6%). Some participants (31%) liked the strong connection between text and picture. 19% participants found INFOGRAPHIC easier for spatial relations and 17% mentioned understanding time was harder. 28% found a clear reading order lacking.

ILLUSTRATEDTEXT—Some participants liked the cleanness and familiarity of ILLUSTRATEDTEXT (11%) and reported that they would use the text to understand the story, and could look up information on demand. However, the majority found jumping between text and picture negative (42%) as they had to create their own connections (14%), complaining about the high density of the text (31%) (*"Along with the high density of text, and hard time of bridging verbal and visual"*).

4.5 In-The-Wild Study

The in-the-wild study was carried out during an international art festival where visitors varied in age, interest, and cultural origin, enabling us to study a more diverse set of participants compared to using standard mailing list recruiting. The study focused on the attraction of each format, how people engaged with each format and for how long, and their preferences for consuming information. To that end, we observed and coded people's behavior as well as conducting semi-structured interviews with some of the readers.

Techniques—We only used INFOGRAPHIC and COMIC in this study. Two comics were created from existing infographics, using the same process as the lab study. Presented topics were chosen for public interest, but unrelated to any of the surrounding exhibitions to mitigate potential confounds on the respective audience attracted by our visualization. To increase attractiveness of the material for the public space, we were more free with visual presentation. Comics were drawn by hand by an experienced comic artist Matteo Farinella. The basic visual encodings in the infographic such as colors, height and size for data variations were preserved, while also using metaphors and sketch styles from data comics (Figure 4.8). We used material

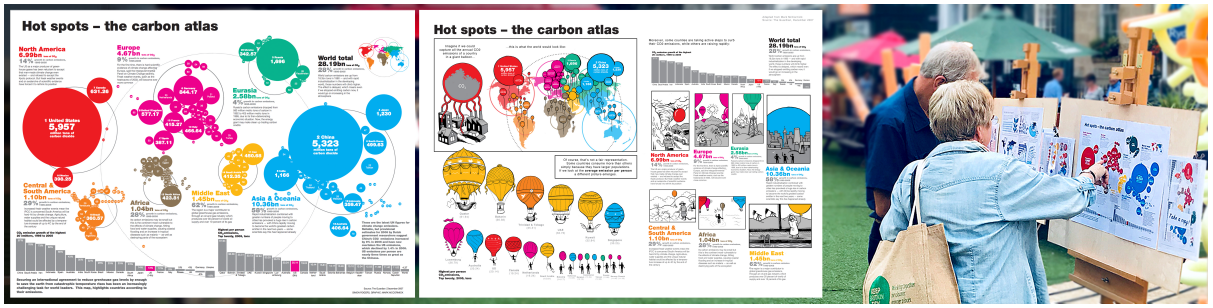


Figure 4.8: Two stories were presented in the wild in both INFOGRAPHIC and COMIC form, for an empirical observation study conducted with pedestrians, measuring reading time, interactions (i.e. pointing and talking) and opinions as evidence for engagement and enjoyment.

on two subjects:

- **Hot spots—The Carbon Atlas (CARBONFOOTPRINT)** [Rogers and Mccormick, 2010] shows global carbon dioxide emission by country, using bubbles of different size to indicate the amount of produced carbon dioxide, colored by region.
- **The Global Water Print (WATERUSAGE)** [Infrastructure, 2010] indicates the volume of water needed for production and human services, showing of amount of freshwater available, the highest consumption, renewable water sources and the highest usage in food production.

Study Design—An A0 sized copy of an INFOGRAPHIC about CARBONFOOTPRINT and a COMIC about WATERUSAGE were placed on a pair of easels in the street, with take-away copies of the comics and INFOGRAPHIC attached (Figure 4.8). After 1h, we exchanged the respective material to show an INFOGRAPHIC about WATERUSAGE and a COMIC about CARBONFOOTPRINT. The study was conducted twice on the same Saturday, once between 11:30am-1:30pm and again from 2pm-4pm. Two study instructors were seated approx. 10 meters distant from the easels, tracking audience behavior. Visitors who read both formats were approached for an interview after they turned away from the graphics.

Data Collection—The time visitors spent at each format was measured, with talking and pointing interactions counted manually. When we interviewed visitors, we asked four questions: why did they stop? Did they finish reading? Was there anything they did not understand? and Which format they found easiest to understand?

4.5.1 Results

During the 4 hours a total of 43 groups stopped for more than 10 seconds to view the graphics. Most **visitors** came in groups of 2 to 4 with a wide age range from adolescent to elderly. From

this, we interviewed 8 groups including 14 people of which 11 were adults, 2 elders and 1 adolescent.

We did not find any differences in time spent on either format or story. CARBONFOOTPRINT comic was read 13 times while Infographic was read 15 times; For WATERUSAGE , both the comic and infographic format was read 14 times. Viewers spend between 8 and 132 seconds reading each story, with averages ranging between 37 and 48 seconds across formats. We did not find a significant difference in reading time.

Counting visitors **interactions** with the material through pointing and discussing each format, we found slightly more engagement with INFOGRAPHIC (8 groups for COMIC vs. 12 groups for INFOGRAPHIC). This could be seen as an indicator that the more exploratory nature of infographics prompts people wanting to share and discuss their observations, while comics focus the reader on understanding the message.

Asked about what attracted them to our graphics in the first place, 4 groups replied *topic*, 1 replied *color*, 2 replied *illustration* in the comic, 1 replied *data*. Asked about which format provided **better understanding** of the content, 6 groups preferred COMIC, 1 group preferred INFOGRAPHIC and 1 group didn't find any difference. Reasons for preferring comics the way it visualized, the grouped messages, the metaphor, and the easy-to-follow layout. Consequently, 12 people reported that COMIC were their overall choice, while 1 adult and one adolescent from different groups would chose INFOGRAPHIC (WATERUSAGE). This supports the evidence from the Lab Study results, that readers find comics more "fun and enjoyable" even if this may not necessarily translate to longer reading time or improved memorability.

Participants took away slightly more copies of CARBONFOOTPRINTin comic format compared to infographics format (20 vs 16). However, this is not the case for WATERUSAGE, for which participants took 16 copies of each.

4.6 Discussion

4.6.1 Main Study Findings

Data Comics improve understanding and engagement—In general, data comics led to more correct answers on average. Data comics have been rated more engaging and more enjoyable, more easy to stay focused, and received better overall ratings. The reasons for these results may be explained by a variety of factors. For example, clear sequencing increases the readers ability to focus and navigate spatial-temporal information, while panels help to divide information into easily memorable chunks, with rows grouping individual messages into higher level messages.

From analyzing specific questions, we found that comics performed better for some information that was explicitly shown and highlighted in panels, such as some temporal events (ALLIANCES) or distributions (ENERGY). On the contrary, not highlighting important information in comics, such as visual encoding (ALLIANCES) can lead to participants overlooking details. These results suggest that increased text-picture integration and more reading guidance (c.f. the design space in Section 3) can lead to better understanding.

Large text-picture distance impairs understanding and increases cognitive load—While illustrated texts were seen as clean and simple, some participants complained about the higher cognitive load required by the constant switching between text and figures. This can explain why participants preferred formats minimizing that distance.

Infographics foster exploration and overview—Infographics are well suited to represent spatial content and are good at delivering both overview and detail. Participants liked the way they allow for comparison, and were more likely to want to share their discoveries with other viewers. We believe techniques from data comics and infographics can be seamlessly integrated with each other, depending on data and message.

4.6.2 Designing Data Comics

Our results can be used to discuss and inform the design of data comic (e.g., [Wang et al., 2019b]). Below, we illustrate some of the complexity of designing good data comics, often requiring multiple trade-offs.

Balancing repetition and highlighting—while most of our results point to an increase in understanding with comics, subjective feedback highlighted potential problems with excessive repetition and sequencing. Too much (visual) repetition and redundancy between panels can lead to confusion as readers struggle to notice the differences. Possible solutions include explicitly highlighting changes, using a cut-out pattern [Bach et al., 2018c] to emphasize small changes, or combining several messages into one panel by using elements from infographics such as annotations. Complicated information could be explained in a large-picture pattern [Bach et al., 2018c] to serve as a mental map, before individual changes are explained in detail.

Balancing sequence and overview—Sequences support temporal and complex causal information, while overviews support comparison and spatial (non-temporal) information and help readers to keep a mental map. The lack of overview has been criticized in data comics, especially if panels show details of the general visualization (e.g., map, scatter plot). However, repetition can be distracting, as mentioned above. A solution could be to carefully pace overview pictures and to make sure zoomed-in content is understood within the larger context. Where

necessary, larger pictures (especially for spatial and detailed visualizations) can incorporate elements from infographics, such as annotations.

Using the layout to structure information—A comic layout provides several means to visually structure information and the story, using panels, panels inside panels, rows and potentially pages. Panel size, number and layout [Bach et al., 2018c] can be used to group and relate messages, to pace reading and attention, as well as to demonstrate importance. A clear page layout, potentially including overview panels, can support information look-up and relation during reading.

Reducing visual complexity—Comics can quickly become visually overwhelming when seen at a glance, as mentioned by some participants. While we designed our comics with this issue in mind, panels full of abstract information remain a natural source of visual clutter, especially if small. Possible solutions include creating larger panels (and hence less panels per page) when panel content gets visually complex. Consistency and repetition of visual information [Qu and Hullman, 2018] can be another solution to keep the overall visual clutter low, if the respective changes between panels are highlighted properly, as mentioned above. Yet, we could not confirm that visual complexity at a first glance actually impacts participants performance negatively.

4.6.3 Limitations and Future Work

Type of stories and visualization—Clearly, the type of story and visualization presented in the studies may influence the the reading experience. We chose maps, networks and scatter plots as representative examples of visualizations.

Future studies need to evaluate whether our findings hold for other visualization types including simpler (e.g. bar charts, line charts) as well as more complicated and less familiar visualizations such as parallel coordinate plots, matrices or tree maps. Such types of visualizations require careful explanations to be understood and used in a storytelling context. We further believe that comics could be successful in achieving this, given their sequential nature and tight integration of textual and pictorial information and we see significant potential for data comics to explain complex data as well as contribute to aspects of visualization and data literacy.

Style and design choices—In creating panels in our data comics, we made specific design choices. Different sequences and pacing could potentially lead to different results [Hullman et al., 2013b] and further studies are required, e.g., in order to determine the appropriate pacing or the amount of redundancy between text and visualization. Similarly, comics may adopt different drawing styles and visualization strategies. For our lab-study, we decided on a simple,

neutral style that used the same visuals and colors of the infographics, to allow for a direct comparison. For the in-the-wild study, we opted for a more visually elaborate presentation to catch people's attention on the street. Hence, INFOGRAPHIC and COMIC were slightly more different in their visual appeal (Figure 4.8). However, interviews revealed that visitors were attracted mainly by the presented topics (CARBONFOOTPRINT, WATERUSAGE) rather than the visuals. It still is possible that introducing more elaborate drawings, characters and metaphors, will affect readers' engagement and attention [Farinella, 2018].

Context and Audience—Our results are naturally limited by the study context and audience. While the audience of our in-the-wild study included a wide range of ages, interests, cultures, and pre-knowledge about visualizations and the presented topics, people may have been reticent to engage with the material at all, whether due to lack of interest or external distraction. Feedback and insights into people's behaviour may also vary in other contexts (e.g., students focused on studying with textbook) but our setup appears to align with the general public's consumption of infographics, as echoed by online news article reading behaviors such as reported in Amanda Cox' talk at IEEE VIS 2011 about consumption of New York Times online articles.[1]

Story Formats—Our design space in Section 4.2 was chosen to motivate and structure our study. For both axes different solutions are possible and there is no unique measure that locates a specific solution. In designing our comics, we made certain choices in panel layout, sizes, style, message chunking, highlighting, text and picture redundancy, which might have had an impact on our results. Moreover, in our first experiment (Section 4.3), story and texts were the same across techniques, in real settings one could be more specific and add more text in infographics. Eventually, as pointed out by Bach et al. [2018c] data comics and infographics span a continuum with terminological and conceptual boundaries not clearly defined. Given our design space, we can adopt the same argument of fluidity for illustrated texts [Bach et al., 2017a]. For example, text plus pictures can include several pictures, pictures can be linked to places inside the text, infographics can contain several pictures, can have more or less linearity, even involve features from comics. Thus, rather than comparing prototype formats, we compared *locations* in our design space. Our study aims to provide some clarity about the usage and impact of each dimension on understanding and engagement. Future designs should take the best of both worlds, i.e., using sequential, narrative, and metaphorical elements from comics, combined with exploratory and image-focused elements from infographics. Gaining more insights into design decisions will be the major challenge for future studies.

Finally, infographics, illustrated texts and comics can be compared along other dimensions and for different contexts: picture size, type of visualization and information, audience and

[1] http://ieeevis.org/year/2011/keynote/visweek/how-editing-and-design-changes-news-graphics

etc [Hullman and Diakopoulos, 2011]. Eventually, comparison with other formats for data-driven storytelling (videos, interactives, physicalizations, etc. [Kraft et al., 2016]) can yield more insights in the respective drawbacks and merits of each format. Better understanding of the potentials of data comics will lead to better authoring support and education.

4.7 Conclusion

The new genre of data comics combines many features with the potential of making data-driven stories accessible and understandable. In order to verify the comics' effectiveness on readers' understanding, memorability, and engagement, we conducted two experiments comparing data comics with infographics and illustrated texts. The findings from this study answer the first research question of my PhD — "Are data comics more effective compared to other data-driven storytelling formats?" Our results are encouraging for the use of data comics, especially for complex spatio-temporal data, which are naturally hard to visualize in infographics.

The results also lead to valuable implications for designing future comics. For example, minimizing the distance between text and its corresponding picture; avoiding too much repetition and redundancy between panels; making sure zoomed-in (visual) content is understood and connected within the larger context; making the comic layout to organize the information. These guidelines complement the use of data comic design patterns, in particular, how textual and visual content placement can influence communication.

By creating study materials in the lab and in the wild, we found creating compelling data comics is not simply integrating texts and pictures in panels, besides the design trade-offs pointed in Section 4.6.2, the comic creator also needs to consider audiences, their data, and visualization literacy, interests and surroundings. So, what are the steps in a workflow of creating a data comic? What are the challenges one may encounter when creating data comics? The study in this chapter provides us primary knowledge towards understanding data comics and motivates us to explore the creation of this format in the following chapters.

Chapter 5

Teaching Data Visualization and Storytelling with Data Comic Workshops

5.1 Motivation

Chapter 4 suggests that data comics leverage the visual language and storytelling concepts from traditional comics (panels, layout, and transitions) to effectively communicate insights from data. In particular, the linear reading sequence and high-level integration of texts and images foster comprehension and engagement. Effective communication with data comics comes down to more elements and factors (e.g., elements featured on the five layers of the Data Comic Creation Model in chapter 3). Creating data-driven stories [Riche et al., 2018] is a labor-intensive process that potentially follows many different paths and stages [McKenna et al., 2017a]. Moreover, data comics require skills and expertise from multiple disciplines such as data analysis for data exploration, visualization design, techniques for data-driven storytelling (e.g., [Bach et al., 2018b]), storytelling in general, writing, as well as comic creation and drawing.

To explore this gap, we designed and ran a workshop on data comics, organized by an interdisciplinary team with expertise in data visualization, graphic design, data comics, and illustration. The workshop presented in this chapter was part of a 3-week assignment in an undergraduate illustration class at the College of Art, Edinburgh University. The workshop aimed to provide students with experience in employing data-driven storytelling and human-centered design thinking. All of the students were asked to document and self-reflect on the design process from beginning to end to enhance their learning as well as act as feedback for the workshop itself. This chapter collaborated with other two co-authors who are experienced lectures in data visualization and illustration, the author of this thesis is involved in the whole process including preparing the slides, conducting workshops, and analyzing the results.

The original publication of this chapter collaborates with Harvey Dingwall and Benjamin Bach [Wang et al., 2019b]. The data comics presented in this chapter were created by the students from the course organized by Harvey Dingwall. The author of this thesis took part in preparing the workshop plan (i.e., procedure, activities, and materials); conducted the workshop as one of the instructors; analyzed the results, and took part in writing the paper.

5.2 Background and Challenges

The design for this workshop was informed by two preceding workshops on data comics at the University of Edinburgh:

Workshop 1—The very first workshop was organized to validate the use of our design patterns for data comics [Bach et al., 2018c]. It lasted for 3 hours with 23 participants from various backgrounds (computer science, design). The workshop was structured into a sketching phase and a story-boarding phase. Students started with a curated collection of material (`https://datacomics925658343.wordpress.com/illustration-workshop`) for various topics including both raw data, sets of visualizations, and data stories in the format of videos and news articles.

Workshop 2—A second workshop was part of a 3-month collaboration between one of the authors and four comic artists and illustrators. The workshop set out with a brief introduction into visualization sketching and design patterns. Each invited artist worked on a specific data set with a respective external data collaborator. During the following months, we had individual check-ins with each group to discuss story, messages, visualizations, and design patterns. To help streamline the individual stories, we asked each group to define three types of messages:

1. **Take-home messages** were high-level messages that we want the audience to remember from the story, e.g., to change their behaviour.

2. **Insights** were facts to support the take-home message(s), communicated through data visualizations.

3. **Data-literacy messages** described general information necessary to understand visualizations and insights: medians, distributions, interpreting error-bars, or specific data transformations. This largely helped constructing the respective data stories.

Both prior workshops borrowed methodology from several related workshops [Huron et al., 2017]. In these workshops, we found that participants appreciated the warming-up sketching practice and the data comics design patterns [Bach et al., 2018c]. The provided data was

appreciated but required significant time for participants to understand (Ch1). We found that participants had problems creating compelling stories and wanted more guidance on this (Ch2). Though we conducted a visualization sketching session at the beginning, participants new to visualizations still had some troubles using purposeful visualizations in their stories (Ch3). Designing compelling layouts that reflect the story-flow was not covered enough through the design patterns (Ch4).

The new workshop, described in this chapter, was part of a 3-week assignment in an undergraduate course for illustration at the College of Art, University of Edinburgh. The course comprised 23 students, many of which had made classic comics prior to the workshop and experimented with variations of the design patterns. A few had also worked with data visualizations but more had experience of info-graphics. The students were confident in drawing so sketching and development was efficient though the challenge in balancing personal styles of drawing and comics with the audience and message of data comics was both one of the key challenges as well as the creatively stimulating elements of the workshop and the final data comics. The assignment asked each student to create one data comic on a subject of their choice using data collected or found by themselves. The comic size was set A3 landscape, allowing comics to be displayed both on screen and in print. The assignment included three 3-hour sessions and a final presentation where students were asked to reflect on the process and results with feedback from peers and tutors.

5.3 Workshop Design and Implementation

5.3.1 Session 1, Week1: Fast-Forward

To quickly get into the concept of data comics and to provide students with an idea of all the steps involved in creating one, we designed a *fast-forward* data-comic workshop in session 1 in which students, in pairs, generated rough data comic storyboards. The fast-forward included 6 phases and lasted for three hours. Slides are linked online (`https://goo.gl/Xz5w43`).

Introduction of data, messages and visualization (30min): The goal of this phase was to introduce students to data visualization, with a strong focus on clarity and communication. We showed a spreadsheet and scatter plot to explain how to interpret data and tell stories about findings. Then we presented problems in effective communication of data visualization done wrongly (i.e. unreasonable design, improperly used visual variables and complex visual literacy).

Introduction to data comics (20min): Data comics were introduced as an approach to improve comprehension of visualizations and the clarity of a message. We then showed examples

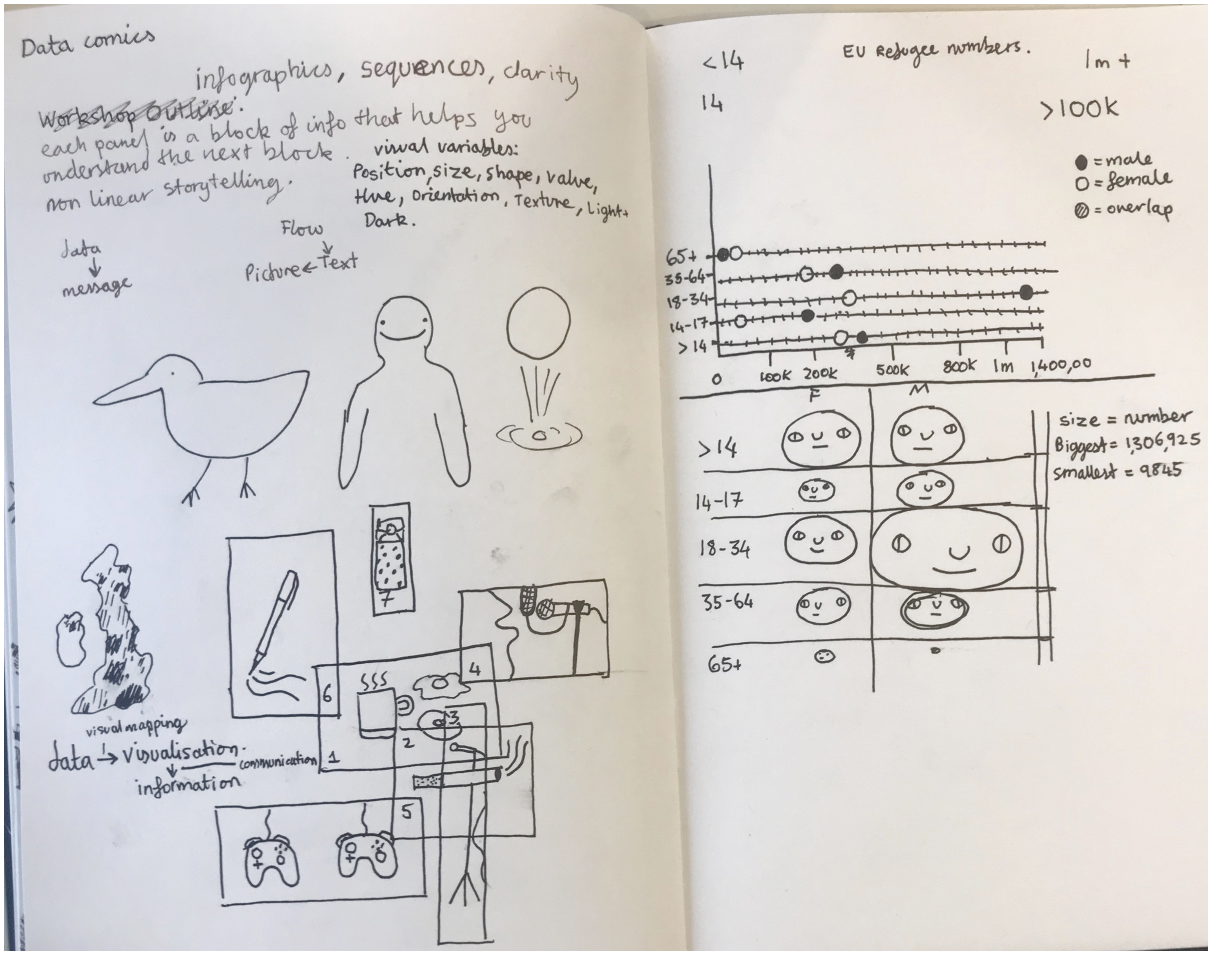


Figure 5.1: Practice in the sketching phase from one of the students.

of data comics for education, personal stories and journalism (`https://datacomics.github.io/`). After initiating the benefits that data comics bring for audiences, we present the features of sequence, presenting time and layout inspired by traditional comics. We then moved to the process of data comic creating.

Sketching (30min): The goal of this session was to warm up students and introduce them to sketching data. The warm-up phase started with some simple 3-second sketches (e.g., draw a bird, a human), then switched to a visualizing task: draw your home country (5sec), then draw the distribution of the population in that country (10sec). Eventually, we asked students to draw 7 random rectangles on a new sheet of paper. We asked students to number each panel in a logical sequence. This first 'comic' layout was the base for drawing the story of their past Sunday in those 7 sequential rectangles (left side in Figure 5.1). After this warm-up, we briefly introduced the notion of visual variables and the idea of visual mapping. To facilitate sketching data (Ch1), we provided students with predefined small data samples and story contexts, used in a workshop on physical data visualization [Huron et al., 2017]. We asked each student pair to

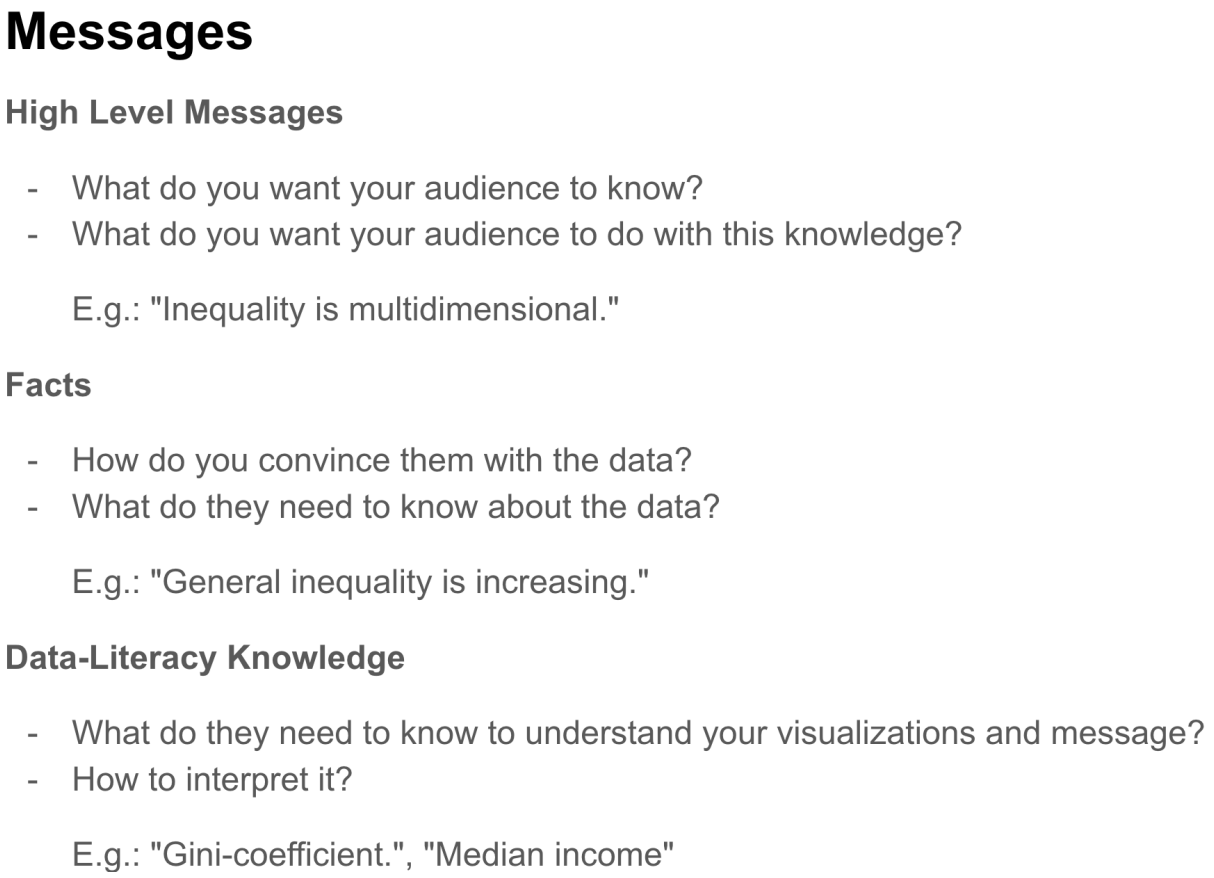


Figure 5.2: This slide displayed a guidance for extracting messages from high to low level, and emphasized interpretation of visual literacy for the visualization to be used.

sketch 3 different visualization designs (3min each, right side in Figure 5.1). At the end of this session, we gave students a 15min break.

Creating data story (40min): Scenario cards [Huron et al., 2017] were handed out randomly, and we displayed the list of messages (high-level, facts, data-literacy) identified in the previous workshop (Figure 5.2). Each card described a context (street, office, etc.) and a persona (friend, peer, etc.) We also presented a story structure example (beginning/middle/end), and a data story example in the form of a video *Wealth Inequality in America* (`https://www.youtube.com/watch?v=QPKKQnijnsM`). Students had 30min to create the story and then 10min to tell their story to their neighbor for self-reflection and feedback.

Storyboard sketch (20min): We introduced how panels work in comics, examples of design patterns and its usage in existing data comics. We suggested the method of using sticky post to structure the panels. Students were asked to draw a draft of the data comic on 1 to 2 A3-sized papers.

Exhibition and presentation (10min): Comics were stuck on a wall and students walked

Figure 5.3: 3-minute storytelling in group and get reflection.

everyone through their drafts while getting feedback from the instructors.

5.3.2 Session 2, Week 2: Working with Own Data

For this 2nd session, students were asked to collect or search for data according to their own interests, as we believed a more personal connection to their data helps telling stories. This reminded students of the process outlined in the fast-forward (session 1) and aimed to yield a first comic draft for each student's individual comic. Slides for this session can be found online (`https://goo.gl/vRZ1Af`). For those students who came with complicated data, we introduced RAWGraph [Mauri et al., 2017] for 10min, visualizing some examples from their website.

Story creating (140min): We reminded students to create a *persona* that would define the target audience of their comic. Then, we went through the same phases as in the fast-forward session; sketching, 3 types of messages, 3-minute storytelling in groups of 3 or 4 (Figure 5.3), drafting layout. We provided students with printed cards of our design patterns [Bach et al., 2018c] and students worked on the first rough draft of their data comic, followed by a walk

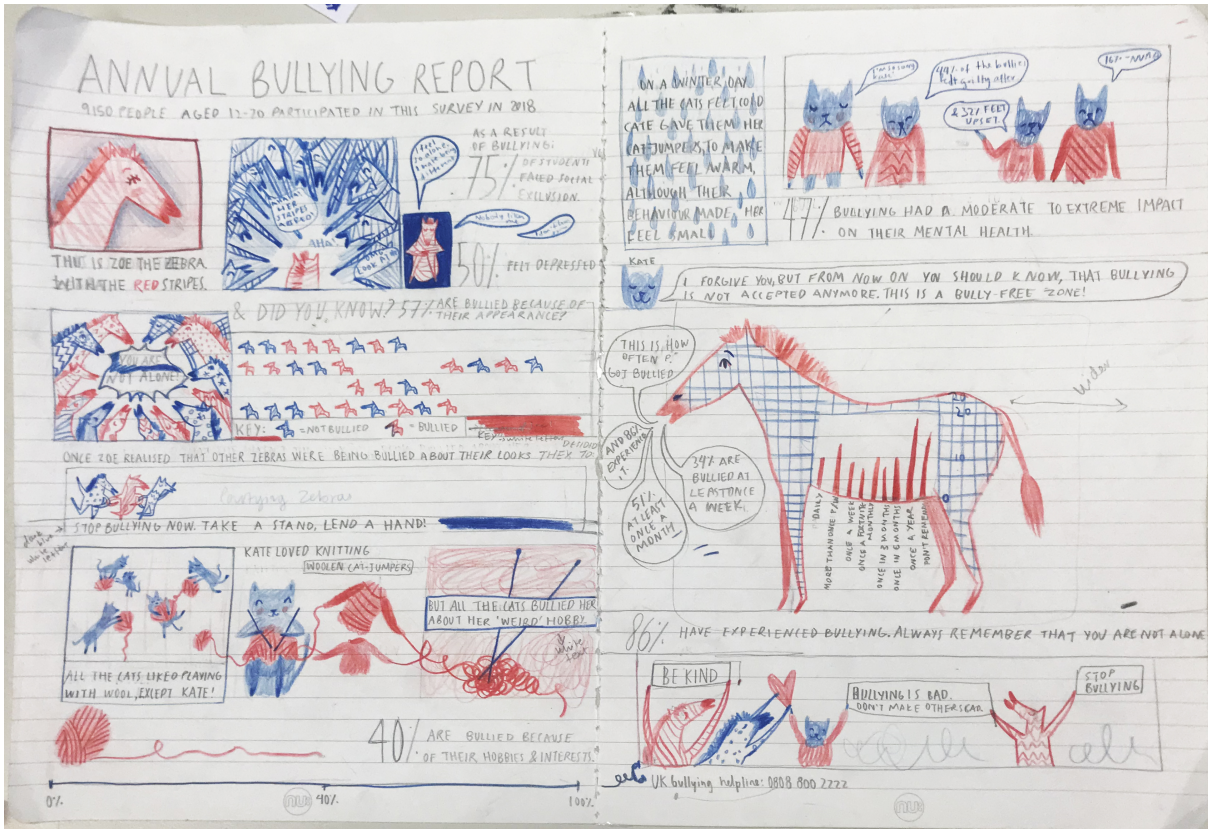


Figure 5.4: *Annual bullying report* (data comic draft) presented in the group-crits.

around the room to look and share their work.

5.3.3 Session 3, Week 2: Group-meetings

Group-meetings were held three days after the second session with groups of 3-5 students. After students presented their drafts individually, instructors and peers offered suggestions for improvement with respect to stories, visualizations, and layout.

5.4 Results

Results have been surprising, surpassing our expectations we had from previous workshops. A final gallery of workshop results is online (`https://datacomics.github.io/`, e.g., Figure 5.9). Data chosen by the students exhibited a great variety of subjects: environmental problems, social phenomenon and health related issues such as bullying (Figure 5.4). Comics used a wide range of design styles where the choices of drawing materials from digital to colouring pencil and use of caricature and pattern added to the richness of the story telling.

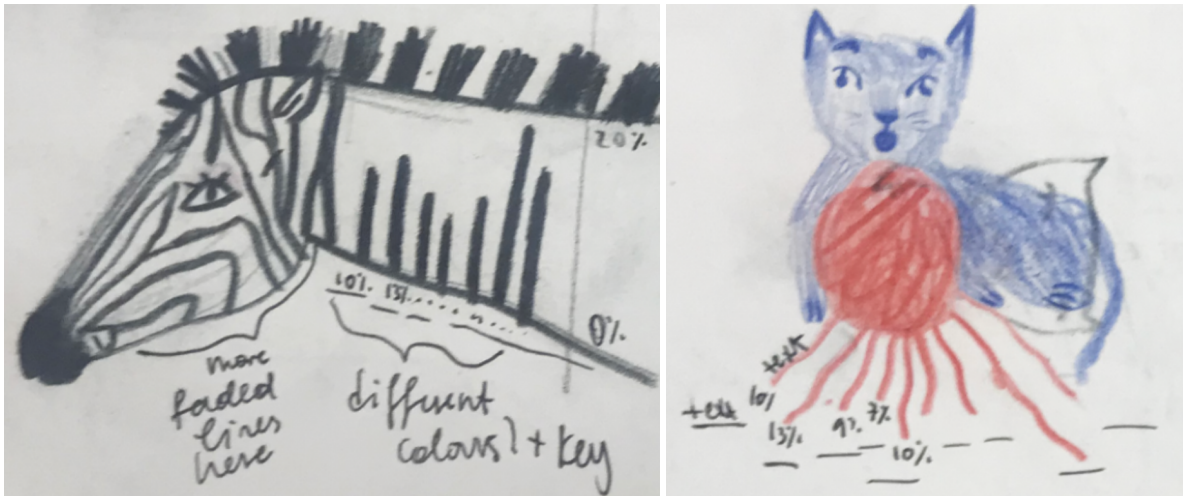


Figure 5.5: Student explored innovative ways of visualizing in the sketch book.

Students' personal drawing methods and interests influenced the style of the comics as much as the nature of the subject matter. Students sometimes blended fact and fiction to emphasize the story; clarify the data visualizations, e.g.'Sustainable Steve' takes us through data on recycling cups and a frog takes us through Mass Extinction (Figure 5.8). Many students used data-comics design patterns [Bach et al., 2018c] including *Question and Answer*, *Multiple Facets*, *Visualization Build-up*, *Zoom*, *Multiple Facets* (e.g., Figure 5.6). Some students created visualization designs themselves, e.g., using stripes on a zebra and threads on a ball of yarn as bar charts (Figure 5.5) or creating cocktail-glass glyphs to report on ingredients in soft-drinks (Figure 5.7). In summary, the success of the final work is relative to the function and context of the data comics but the importance of the workshop was in learning for the individual students. The diversity of research, processes and skills involved in making data comics made for a particularly rich learning experience. The full list of results can be found online: `https://datacomics925658343.wordpress.com/illustration-workshop`.

5.5 Discussion

The following discussion aims at reflecting on future improvements and exercises that can further improve the learning of drawing data comics. Our reflections and interpretations are naturally constrained by the format of the workshop we ran. Specific challenges might not arise for different backgrounds of participants and more time available. Most reflections stem from direct observations, discussions with students, and a brief formal questionnaire.

Creating stories required students moving back and forth between the story, visualizations,

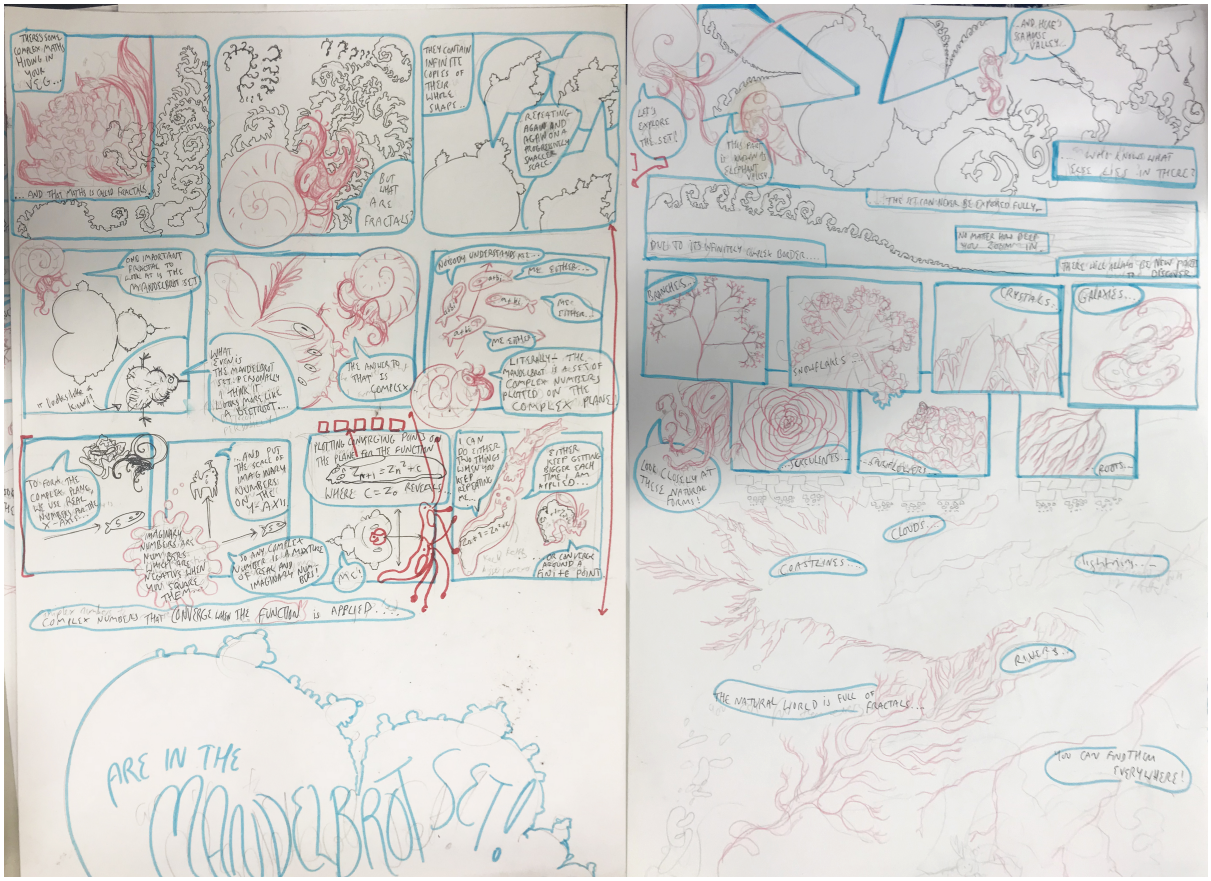


Figure 5.6: *Mandelbrot set.* The author used the *Question & Answer* and *Visualization Build-up* to explain the mathematical principles with a coordinate system, used *Zoom* and *Multiple Facets* to engage audience to explore its richness.

and the data. During story creation (Session 2), some students realized they required more data to provide additional evidence or information from other aspects. We believe this is an important message to teach/learn that data collecting can happen at the story creation process. Also, students could effectively organize the layout, develop clear reading sequences and apply design patterns. Personas and scenarios helped students set their language tone or comic style and make informed design decisions (e.g. color usage and characters). We also conclude that the fast-forward (session 1) was very useful and students found the provided data helpful. From the interviews and observations, we obtained further insights into challenges and suggestions of potential solutions.

Tutorial material: In our first workshop, we started from visualizations, data sets, or stories on specific topics (immigration, CO2 pollution, inequality, etc). As we decided to run a fast-forward workshop this time, we wanted to rely on similar data that was easier to grasp and to work with. While still considered useful by students, we found two problems with this: students

Figure 5.7: Student explored using different variables to present the proportion of calories, sugar and juice concentrate in drinks.

struggled coming up with good 'stories' in the sense of describing the data and report simple insights. The reasons may have been the nature of the data itself, but more likely the missing personal relationship to the data. In other words, we believe it is crucial for storytelling to a) choose the data set one wants to work with (other than randomly getting data assigned) and b) choosing data that matches personal interest and knowledge to connect and comprehend audience and message. This was evident in the second phase, where students worked with their own data. Deciding which material to start with eventually depends on the participants and their background, as well as on the nature and intended outcomes of the workshop. Ideally, participants come with a good understanding of their data. If this is not possible, providing simple data sets to chose from and giving aids to their understanding is helpful. Extended versions of a data comics workshop should include more time for exploratory data analysis, e.g., using Tableau.

Visualizing data effectively: Most comics involved characters and descriptions of the data context (e.g., Figure 5.8) such as a characterful endangered frog trying to comprehend

and come to terms with the sixth global mass extinction! However, some comics showed a few data visualizations only. In the future, we would put more emphasis on visualizations and visualization design. Perhaps the most compelling challenge for students is to explain visualizations properly. With more time, we need to emphasize that visualizations are complex constructs that often require explicit explanations to guide the audience through an argument rather than illustrating a number.

Creating narration: Some students reported problems with creating stories, which given their background, was not surprising. One obvious solution is to provide more guidance on how to develop and structure a story; experimenting with beginning-middle-end, providing context, describing data, introducing visualizations, report on insights (using visualizations), build a climax, formulate a conclusion and an outlook. Potential material may include narrative design patterns [Bach et al., 2018b]. However, we see a general lack of understanding and appropriate literature in this field.

Creating repeating comic drafts: Some students had problems with creating multiple drafts for their storyboards and layouts. The reasons being the additional workload of redrawing existing panels. A possible solution would be to use movable cards or post-its to draw individual panels onto them—exercises we did not manage to perform as during the workshop we had to attribute more time than expected to certain exercises. Moveable cards can easily be re-arranged and extended in timed exercises. An alternative could be to use tracing paper, placed on top of the previous draft, to aid roughly sketching new drafts. Given the troubles in creating stories, we believe experimenting with the storyboard (sequence of panels) is a crucial feature that will have to occupy significantly more time in future workshops as we are building our knowledge base in critically analyzing data comics.

Working with design patterns: Design patterns aimed to structure the creation process by providing participants with design templates. However, we found it hard for participants to understand patterns in that short amount of time as patterns are learned through (re)application. We had planned, but could not do more specific exercises on the patterns themselves due to time constraints. However, in future iterations of this workshop, we encourage short, yet precise exercises with the patterns and participants own data. For example, “create a *build-up* for your visualization”, “describe change in your data through *annotated transitions*”, or “split your argument into individual panels using *multiple explanations*.”. Exercises could be very brief, e.g. 5min each. Eventually, higher-level narrative patterns [Bach et al., 2018b] could help as well, but would require significantly more time to present and internalize.

Working with data visualizations: Data visualizations play an essential role in data comics for providing data-based evidence. Participants from art and design background have little

Figure 5.8: Examples of using character to guide through the story.

Figure 5.9: A polished example by one of the student.

experience of working with data and visualization techniques. Although we included visual variables, basic visual encoding principles in the introduction, and introduced RAWGraph [Mauri et al., 2017], participants still struggled with visualizing data, and explaining the visualization created by themselves. This challenge had two aspects, first, to learn how to create and use data visualization techniques; second, to sufficiently explain the visualization used in the data comics, which is especially desired for non-trivial visualization techniques such as adjacency matrices, box plots, and visualizations created in creative styles by themselves. Currently, there is no shortcut for learning data visualization, nor guidance for creators to explain visualization techniques in data comics.

5.6 Conclusion

In this chapter, we detailed our methodology for a revised data comics workshop, informed by two previous workshops. Working with illustrators introduced an emphasis on the style and genres of data comics that can be considered in future workshops and research. Our

reflections highlight future improvements and design considerations for data comics workshops and hopefully will pave the way for more workshops and teaching methods for communicating and engaging with data.

This chapter also discussed the challenges of creating data comics with workshop settings, and imply a future research direction — how to support creators who are new to visualization learning and explaining visualization techniques in data comics with limited time?

Chapter 6

Cheat Sheets for Data Visualization Techniques

6.1 Motivation

Data comic workshops provide an opportunity for participants to quickly experience the process of making data comics. The results from Chapter 5 present both richness in contents and expressiveness in visual and narrative presentations. Meanwhile, the workshop implies challenges in using and explaining visualization techniques. This challenge is not only identified with data comic creation. With a rapidly increasing array of techniques and applications across different fields, we see a natural demand for instructional resources to support a growing and heterogeneous audience in creating and decoding data visualizations.

There is no shortage of resources to learn about data visualization, including books (e.g., [Kirk, 2016, Munzner, 2014, Cairo, 2013 - 2013]), survey papers, workshops (e.g., [Roberts et al., 2016]), classes, and online catalogues of techniques [Ribecca, 2019]. However, much of this material talks about general topics in visualization such as pitfalls for misinterpretation, design, presentation, tasks, storytelling, or layouts. We find a lack of material that supports the understanding of specific visualization techniques: *How to decode visualizations correctly? What patterns can we see and how should we interpret them? What are common misinterpretations to avoid for a given technique?*. There is a specific need for resources that are tightly structured and standardized, which align with established visual design principles and which are easy to understand.

In order to support the development of material for learning, teaching, and the regular use of visualization techniques, this chapter presents *cheat sheets for visualization techniques* (Figure 6.1). In particular, we are interested in how to design cheat sheets that are understandable

and usable. Our cheat sheets are inspired by infographics, data comics [Zhao et al., 2015b, Bach et al., 2017a, Zhao et al., 2019], assembly instructions, and a wide range of examples of cheat sheets for programming languages. In the context of visualization techniques, we **define a cheat sheet as** *"a set of concise, annotated graphical explanations of aspects of a specific visualization technique"*. They aim to provide carefully designed visual and textual explanations in a concise way while focusing on the most important knowledge necessary for given tasks. We imagine cheat sheets to support two scenarios: *i)* first-time *learning* aided through slides, posters, books, or activities, and *ii)* as *look-up* material during an actual data exploration process. Our cheat sheets, in their current form, do not aim to overview and classify visualization techniques, nor to support finding the "right" technique for a task.

In this chapter, we present cheat sheets for a selection of non-trivial visualization techniques representative to different data types (temporal, relational, hierarchical, statistical/quantitative, multidimensional) and taught by our collaborators: *parallel coordinates plots* (PCP), *adjacency matrices*, *Whiskers Plots* (boxplots), *tree maps*, *confluence graphs* [Bach et al., 2017b] and *time curves* [Bach et al., 2016b]. Using an iterative design process that involved 8 experts and regular feedback, we designed six different *types* of cheat sheets for each technique, each of which explains a specific aspect of a technique and supports different usage scenarios: (a) *anatomy* explaining visual elements, (b) *construction* explaining the general idea of a technique, (c) its *visual patterns*, (d) common *pitfalls*, (e) *well-known relatives* relating a visualization to a more commonly known representation, and (f) *false-friends* which show visualizations that look similar but have to be interpreted differently. *Introduction* is to give a high-level overview of the technique.

Visualization Cheat Sheets are our approach to support visualization literacy by providing a concise and consistent set of explanations for understanding specific visualization techniques, their visual marks, and visual patterns. As learning is a process, our sheets aim to support a range of tasks and steps related to visualization literacy while focusing on supporting learning and lookup during use.

The original publication of this chapter collaborates with Lovisa Sundin, Dave Murray-Rust, and Benjamin Bach [Wang et al., 2020]. The author of this thesis designed the first three sets of cheat sheets (i.e., *parallel coordinates plots* (PCP), *adjacency matrices*, and *Whiskers Plots* (boxplots)); conducted the expert workshop, expert interview; collected feedback and iterated the cheat sheets; conducted readability evaluation with non-experts; and took part in writing the paper.

6.2 Cheat Sheets and Visual Explanations

The traditional usage of 'cheat sheet' is *"a piece of paper bearing written notes intended to aid one's memory, typically one used surreptitiously in an examination."* [oxf, 2019, Dorsel and Cundiff, 1979]. Using—but more so preparing—cheat sheets has been found to improve students' exam performance [Wachsman et al., 2002, Song and Thuente, 2015], especially when students spend time with preparing cheat sheets [Erbe, 2007, Hsiao and López, 2016]. Suggestions for good cheat sheets include hand-made content and good (visual) organization [Song and Thuente, 2015]. Consequently, cheat sheets have been used to learn programming languages and support their daily use [van Fjodor, 2016, Lewis, 2018, Basawapatna et al., 2010, Hsiao and López, 2016], data analysis [dat, 2019], and types of neuronal networks [van Fjodor, 2016]. These cheat sheets provide the most important information at a glance, and are intended to be looked up during a specific task. In data visualization, types of cheat sheets have been used to help designers choose a chart type [Franconeri, 2018, Smith et al., 2019, Few, 2012] and design effective visualizations [Schwabish, 2019, Iliinsky, 2019]. However, similar to the visualization chooser cards [Evergreen and McMahon, 2019] and flashcards [Schwabish, 2016], these approaches either do not focus on techniques, or concentrate on cataloguing techniques rather than supporting use and understanding.

Our cheat sheets draw on this practice of providing important information at a glance in a structured form, and aim to support the learning of visualization techniques. This is distinct from the process of *choosing* a specific visualization type, instead dealing with the construction and metaphor, the variety of visual patterns, possible visual misinterpretations [Pandey et al., 2015] specific to a particular technique. To provide appropriate visual illustrations, we draw inspiration from infographics and comics. Comics in particular, with their unique combination of complementary text and pictures, sequential narrative, and general familiarity, have been found beneficial for science education [Farinella, 2018, Kim et al., 2017a], communicating complex data-related concepts [Wang et al., 2019c] including confidence intervals, hypothesis tests, regressions, and simple statistical graphs [Klein, 2013, Smith and Gonick, 2005]. More recently, comics have been used as general format for data-driven storytelling medium [Zhao et al., 2015b, Bach et al., 2017a], coming with specific design patterns [Bach et al., 2018c].

6.3 Design Principles

Based on early exploratory prototypes and background literature, we identified the following initial design principles for creating cheat sheets:

- **D1—Modularity:** To introduce structure, we propose *types* of cheat sheets, each focusing on explaining a specific aspect such as visual patterns or visual parts. This is to reduce overcrowding, and to allow for different combinations of sheets to support different situations and different *formats*.
- **D2—Context independence:** Sheets should not rely on specific data examples or external contexts.[1] Our goal is to explore the *content* of cheat sheets and which concise graphical illustrations are required to communicate this content. By keeping our sheets context-free and abstract, we aim for reusability across examples and to learn about how abstract sheets *can* be while still allowing people to apply them to their specific examples at hand.
- **D3—Clear graphics:** As data visualizations are a form of visual languages and aim to facilitate memory and lookup, our goal with cheat sheets is to explain as much as we can through graphical content and use text where graphical content alone is not sufficient.
- **D4—Style neutrality:** Strong graphic styles can have implications and can render a graphics work specific to a narrower audience [Madden, 2005]. Following our principle of context-independence, with this paper, we aim to explore cheat sheets for a general audience and to allow for simple adaptation to more specific cases where necessary. For this work, we used stylistically consistent simple black-white graphics, partly inspired by the very minimalist but clear *xkcd* style [Munroe, 2015], while at the same time avoiding overly minimalist designs [Bateman et al., 2010]. Although color is an important visual variable, we focus on the basic structure and layout of visualization techniques; color is a general design decision whose application is often independent from a specific technique and therefore greatly varying with the specific visualization instance.

6.4 Research and Design Methodology

We designed a collection of cheat sheets using an iterative seven-step design process.

1) Idea generation: we selected several common yet non-trivial visualization techniques such as treemaps, adjacency matrices and horizon graphs, and used these to explore a range of cheat sheet types to explore various aspects: comparison, variation, advantages and drawbacks, concrete examples etc. **2) Expert workshop and co-design:** we conducted a two-hour co-design workshop with three senior university lecturers specialising in geographic visualization, medical informatics, and data visualization. Participants were asked to introduce a visualization technique, i.e., a parallel coordinates plot (PCP) to students unfamiliar with that techniques using

[1] Note that following the Expert Workshop in Section 6.4 we relaxed this principle

sketches and as little text as possible (D3). All participants created material in a comic style with concrete examples and a story to motivate and introduce the specific technique (e,g., peoples' biological characteristics, fruits and nutrition facts). Participants were then asked to feedback on the draft sheets prepared in step 1 (visual patterns, pitfalls and, construction, variations, and comparison) for PCPs, boxplots, and treemaps. The workshop highlighted the need for concrete examples (contrary to our principle of context-independence) and helped us narrow down our choice to the final set of six sheet types. **3) Expert interviews:** after redesigning our sheets based on feedback, we consulted individually with four university lecturers (medical informatics, graphic design, illustration and digital learning), on potential use cases. As well as small design tweaks, suggestions included: handing out the *Introduction* before a lecture and *visual patterns* during and after the lecture; handing out all sheets in a creative course on physicalization to inspire students; adapting the style and data examples from the *Introduction* to the specific context of their own lectures. **4) Readability feedback:** after another set of internal iterations, we conducted individual readability walk-throughs with four graduate students in design and computer science, all of whom had briefly encountered the respective techniques in their lectures. We obtained feedback on ambiguities that required further explanation and how much they have understood the concepts. **5) Design-by-example:** we asked a researcher and designer unfamiliar with our research to create cheat sheets for additional visualizations based on our examples (they are now a coauthor of this chapter). The example's structure was found to be straightforward and salient enough for the new researcher to produce sheets for three additional visualizations with only minimal consultation. The main question involved the *Introduction* sheet (explained in Section 6.5), the structure of which was deliberately left open to encourage tailored story-lines for the specific visualization and the data examples necessary to explain a visualization. **6) Readability study with non-experts:** to obtain further feedback on readability, usability and to refine our design, we conducted an empirical user study following the rapid iterative test and evaluation method [Medlock et al., 2005] with 11 participants novel to visualization (Section 6.7). **7) Guideline workshop:** we eventually, conducted two workshops with a total of five visualization experts and non-experts to apply, evaluate, and refine design guidelines for creating additional cheat sheets.

6.5 Cheat Sheet Types

This section reports on six types of cheat sheets informed by the expert workshops, expert interviews and the readability study with non-experts. According to our design principle of modularity (D1), this section describes cheat sheets for individual aspects of visualization

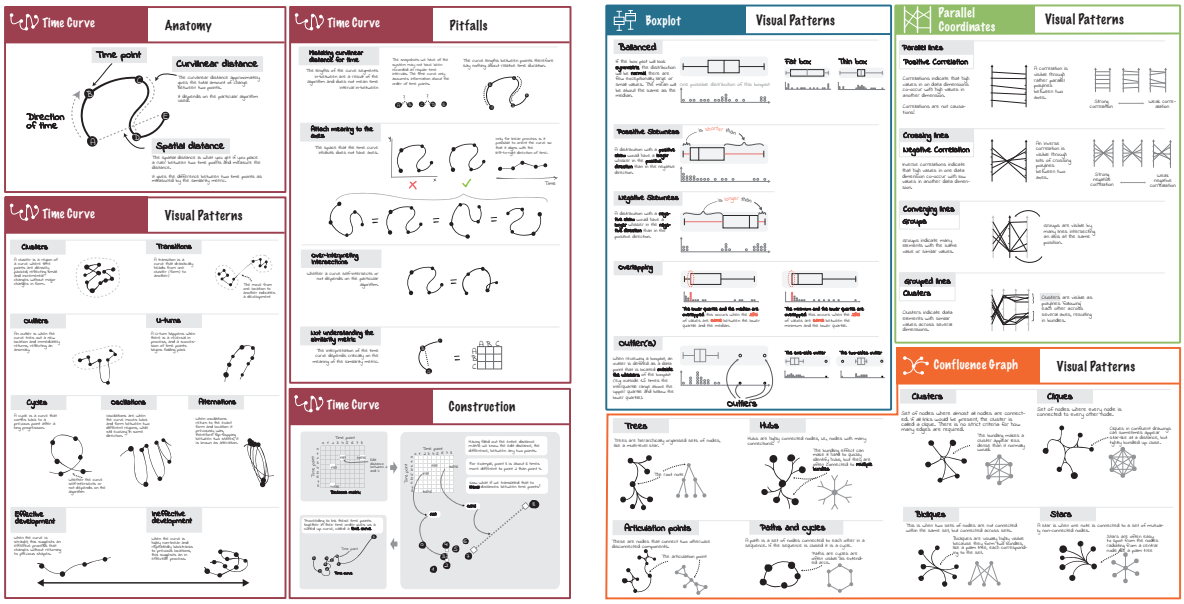


Figure 6.1: Examples of customized cheat sheets combining several aspects of visualizations. Left: example of look-up summary sheets for the time curve visualization showing *anatomy*, *visual patterns*, *construction*, and *pitfalls*. Right: Cheat sheet showing *visual patterns* for several techniques: boxplots, parallel coordinates, and confluent graphs. More cheat sheets and information can be found at: `https://visualizationcheatsheets.github.io`

techniques. Figure 6.1 gives an idea how different sheet types can be combined into a *format* for a final presentation. For each type we give a brief definition, explain its motivation and its design.

ANATOMY *explains the visual elements of a visualization technique, their composition, their specific terminology, and how they relate to the data (Figure 6.2).* The visual components of a visualization can include individual visual marks (in the sense of Bertin et al. [1983]) as well as groups of marks, axes, locations in a visualization etc. Our design for *Anatomy* was inspired by graphics-first explanations in patents using a bare minimum of ink while relying on black, white, and shades of gray only. Text labels are connected to their components through simple leader lines. We found this design to increase readability, communicating on an abstract level without confounding the reader with specific but unimportant details (e.g., thickness of axes, number of lines, etc.).

Often, visualizations use similar visual marks to refer to different parts of a visualization, e.g., lines in a PCP are used for both dimensions and elements. Other visualizations use visual marks with very specific meaning such as the Whiskers in a boxplot. Constructing precise terminology here is a challenge, as *i)* competing terminologies exist ('cell' vs. 'mark' in an

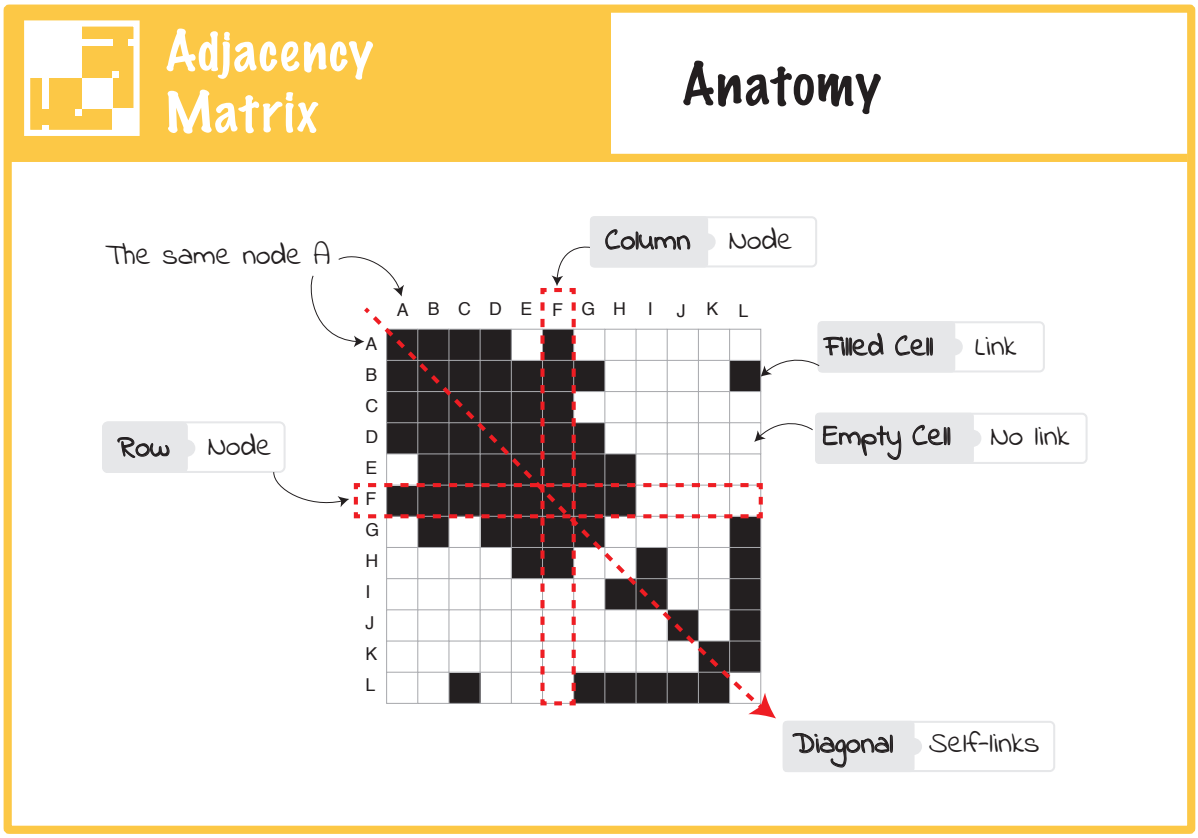


Figure 6.2: *Anatomy* sheet for adjacency matrix; labels for visual elements are shown on a gray background for highlighting while data terms are shown on a white background linked to the visual terminology; arrows are curved to not interfere with the rather geometric visualization; red is used to show information overlaid on black elements.

adjacency matrix); *ii)* some elements do not have defined terminology (the box in a box plot); *iii)* visual representations have distinct names from the data concepts they represent ('axes' as visual representations in PCP versus 'dimensions' as the underlying concept). We show the visual term on a gray background while the data term is shown on white background associated through a jigsaw shape. For simplicity we did not include alternative terminology (e.g., 'node', 'vertex'). Additional explanations are added in sparse textual annotations. Red lines are used whenever too much information is overlaid.

CONSTRUCTION *explains the conceptual idea behind a visualization design and how it encodes data.* It reflects the relation between the data and its visual encoding by demonstrating the process of constructing or providing a familiar metaphor (e.g., Figure 6.3) to help understand and memorize the encoding mechanism. While *Anatomy* explains the visual elements and their terminology, *Construction* provides a procedural explanation of a visualization design and delivers a blue-print for how to explain the visualization to a larger audience. For example,

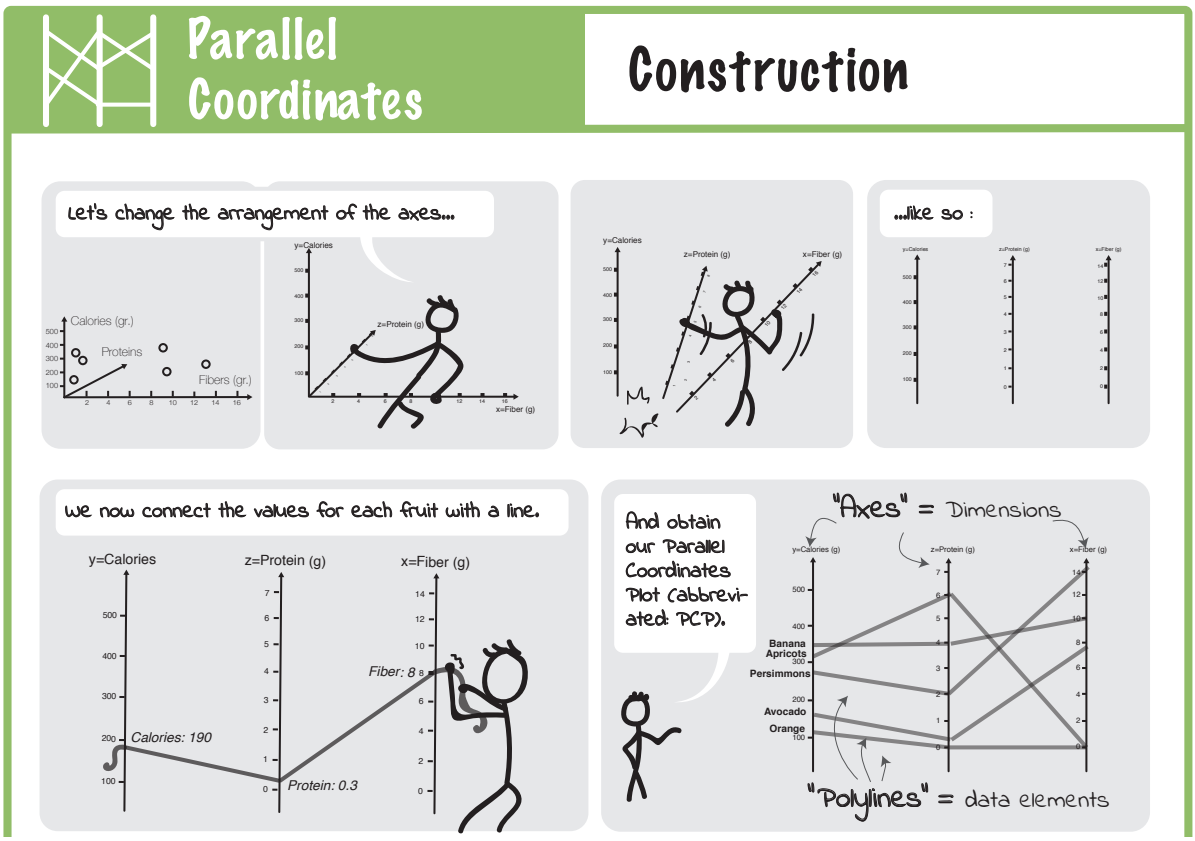


Figure 6.3: Part of Construction for PCP, showing its "creation" from three-dimensional scatterplot in a comic-strip.

in a report, a presentation, or video, a presenter can show and explain a *construction* to the audience *before* talking about results in a visualization. Our constructions were largely inspired by data comics (e.g. the *build-up pattern* [Bach et al., 2018c]) and videos that explain the visual encoding process in animated presentations (e.g., [box, 2019]) as well as explanations in the papers introducing visualization techniques, e.g., [Henry and Fekete, 2006, Bach et al., 2016b, 2017b, Wickham and Stryjewski, 2011])

WELL-KNOWN RELATIVES *highlights other approaches to visualizing the same data.* Relating new knowledge to existing knowledge is fundamental for learning, and there are often complementary techniques for visualising a given dataset to draw on [Chang et al., 2017], e.g., boxplots, barcharts and swarmplots for distributions, scatterplots and PCPs for multidimensional data. Figure 6.4 shows node-link diagrams as a relative of adjacency matrices, inspired by Henry and Fekete [2006]. Node labels and groups are the same in both diagrams, while elements are highlighted and linked either implicitly (red clusters) or implicitly (element highlighted in red and green).

VISUAL PATTERNS *sheets (or "patterns" for simplicity) provides a catalogue of meaningful*

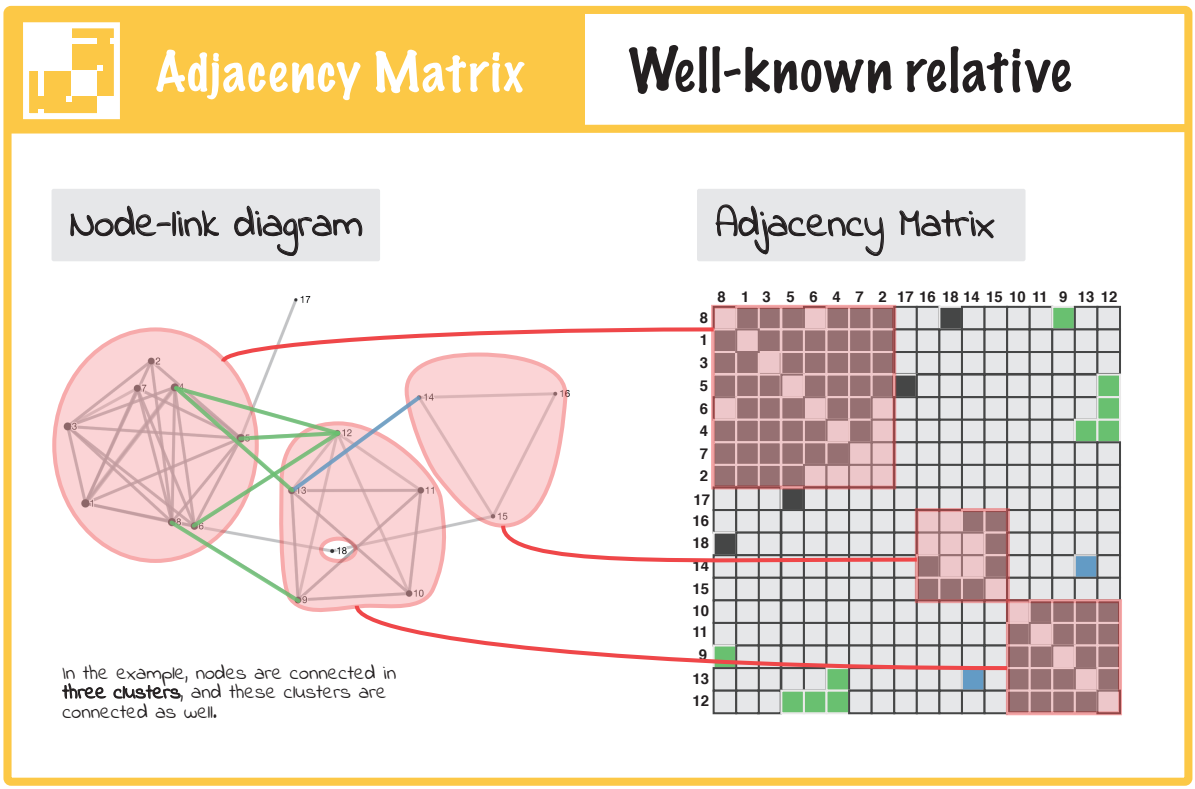


Figure 6.4: Well-known relative for adjacency matrix, adapted from Henry and Fekete [2006]. Node-link diagram are familiar to most people, helping to understand the less familiar matrices: Blue and green represent the connectors between clusters in both charts.

patterns that may occur in a visualization. Spotting patterns can be key to correctly interpreting visualizations and making discoveries—for example, network cliques in adjacency matrices are visible as solid blocks (Figure 6.5, relies on a good ordering [Behrisch et al., 2016]). By providing a set of patterns and their explanations to users, cheat sheets can support users in developing understanding using a visualization.

Many patterns are described in the literature for specific techniques (we draw on Behrisch et al. [2016], Bach et al. [2016b, 2017b], Wickham and Stryjewski [2011] here). Terminology is not always clear, e.g., naming either the visual representation or the concept being represented: 'block' versus 'clique' for networks, 'crossed lines' versus 'inverse correlation' and so on. For our examples, we chose the 4-8 most common patterns for each visualization and describe them as follows:

- **Title** gives the visual appearance ('block') while **subtitle** gives the interpretation ('clique') with a **description**.
- A **visual explanation** that shows an abstract example of a pattern, using annotations and

shading to explain concepts.

- **Variations** of the patterns are used to show differences in size, expression, location etc. From the user study we found *variations* to be essential for understanding patterns and to correctly interpret a pattern.

PITFALLS *show possible misinterpretations of a visualization.* Distinct from ways that visualizations can be poorly designed (e.g., missing labels, different but hard to perceive colors, not colorblind safe, or deceptive titles [Pandey et al., 2015, Carr, 1999, Jones, 2011, Monmonier, 2018, Kong et al., 2018]), pitfalls are specific to a given technique, and can occur even with well designed outputs. We can group pitfalls on our sheets into three major groups: human inattention (e.g., missing to check for min, max value on an axis), visual encoding (e.g., in a boxplot where the minimum and the lower quartile overlap), and data features obscured by particular visualization techniques (e.g., boxplots are not able to show whether a distribution is bi-modal or normal distributed). Our design for the *pitfalls* sheets uses three strategies (Figure 6.6): *i)* highlighting problematic parts of the visualisation (e.g., scales in PCPs do not always start from '0') with arrows, circles or an exclamation mark symbol ⚠; *ii)* using comparison to address the difference (e.g., equal versus not equal in a matrix and ascending scale versus descending scale in a PCP); and *iii)* using '✓' and '✗' for things to do and things to avoid when constructing visualisations (e.g., do not show axes for time curves [Bach et al., 2016b]).

FALSE-FRIENDS *lists visualization techniques that are visually similar but functionally different.* For example, boxplots share visual similarities with candlestick charts and error bars in bar chars (Figure 6.7); parallel coordinates can look similar to line charts, time curves can look similar to connected scatterplots [Haroz et al., 2016]. Our *false-friends* module aims to clarify these differences by placing the 'false-friend' beside the actual visualization, separated by a '≠' sign. For each 'false-friend', we add one sentence to introduce the applications and annotated the different components.

6.6 Formats

Figures 6.1 and 6.8 show *formats* of cheat sheets, i.e., cheat sheet types organized for a specific presentation purpose. While the examples in Figure 6.1 show two ways of grouping cheat sheets—by type (left) and by technique (right), Figure 6.8 shows a more elaborate example. An ***Introduction*** *is a format that* introduces a visualization technique within a specific context, explains its purpose, and highlights important aspects described in the individual sheets. An *Introduction* is a more elaborate component that can include information from other sheets to

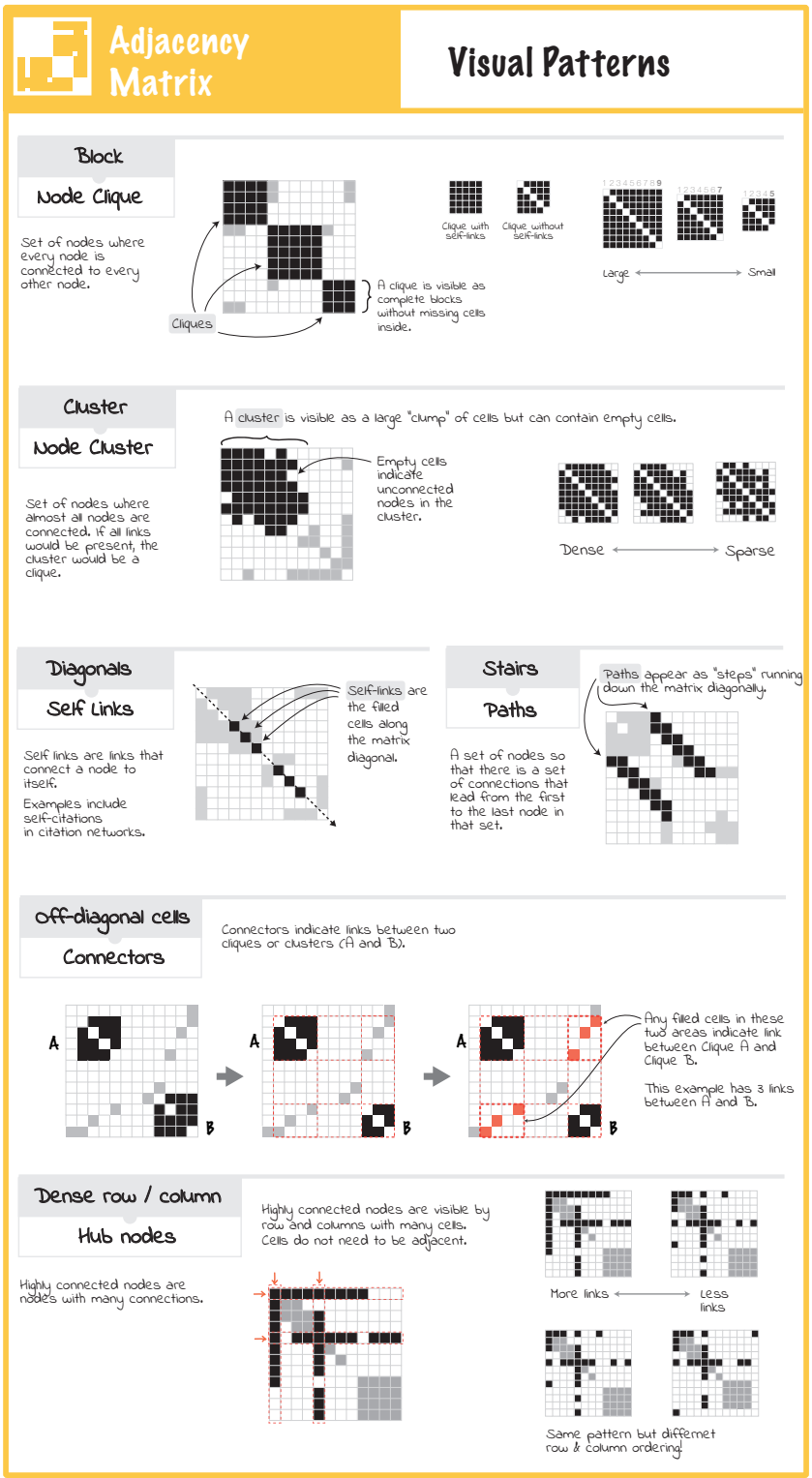


Figure 6.5: *Visual patterns* for adjacency matrix. Patterns are separated from each other by gray dividers. Patterns are given two names, a visual name (e.g., 'Full Block') and a data name (e.g., 'Node clique'). Most pattern are shown by an example, complemented with text. Pattern variations (e.g., dense *vs.* sparse clusters) use scaled-down graphics.

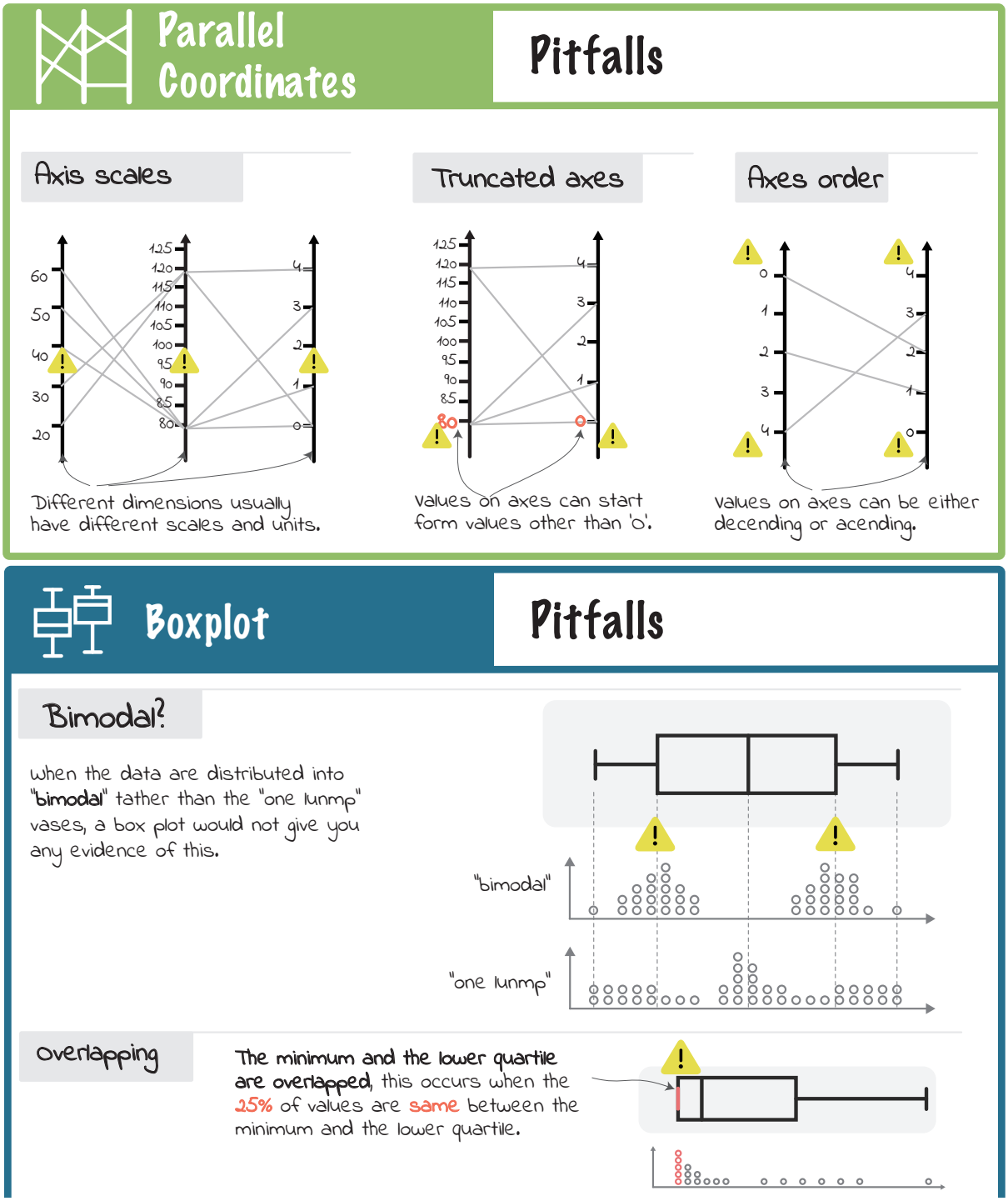


Figure 6.6: Examples of pitfalls for PCP (top) and boxplot (bottom). Yellow exclamation mark signs show areas where a pitfall can occur.

provide an introduction into the technique and overview over information in the related sheets (Figure 6.8). For our *Introduction*, we were mainly inspired by data comics. After several iterations and feedback from our expert workshop, we settled on the following structure for an

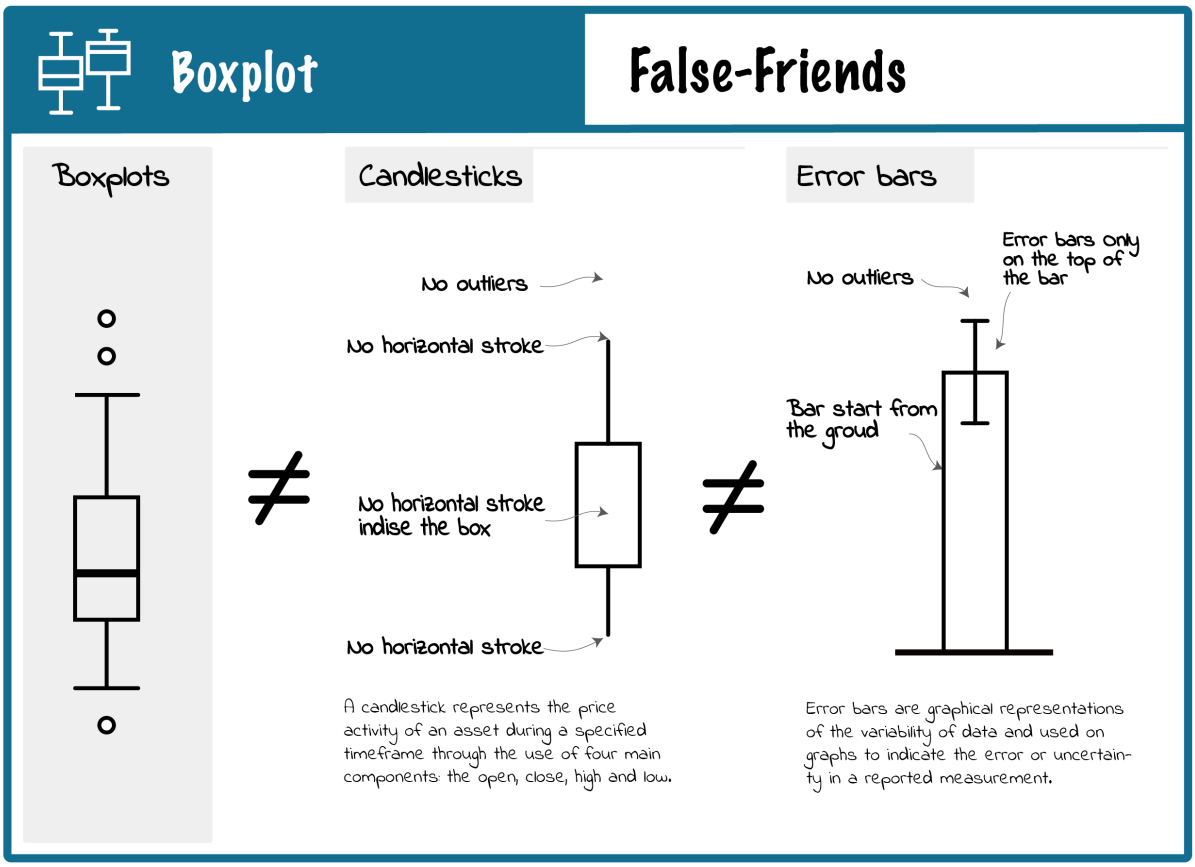


Figure 6.7: *False-Friends* for boxplots: the target visualization is shown on the left on a grey background to separate it from the rest. Next to it, separated by a '≠' sign, are shown visualizations that look similar but have different meanings and uses, annotated with main differences and brief explanations of their use.

Introduction:

1. **Context:** start with a real-world example (e.g., fruits) that introduces the data type (e.g., multivariate data) and high-level tasks (e.g., finding fruits for with balanced nutrition values). This can include characters if necessary.
2. **Well-known relative:** then, introduce a visualization technique familiar to your audience (e.g., node-link diagrams) and explain its limitation to motivate why an alternative techniques is required.
3. **Construction & anatomy:** then, show a *construction* to introduce the novel visualization and explain its visual components with the *anatomy*.
4. **Visual patterns & pitfalls**: explain how to read the visualization by explaining some visual patterns in the context of your data example from step (1). If necessary, mention some pitfalls in the data example.

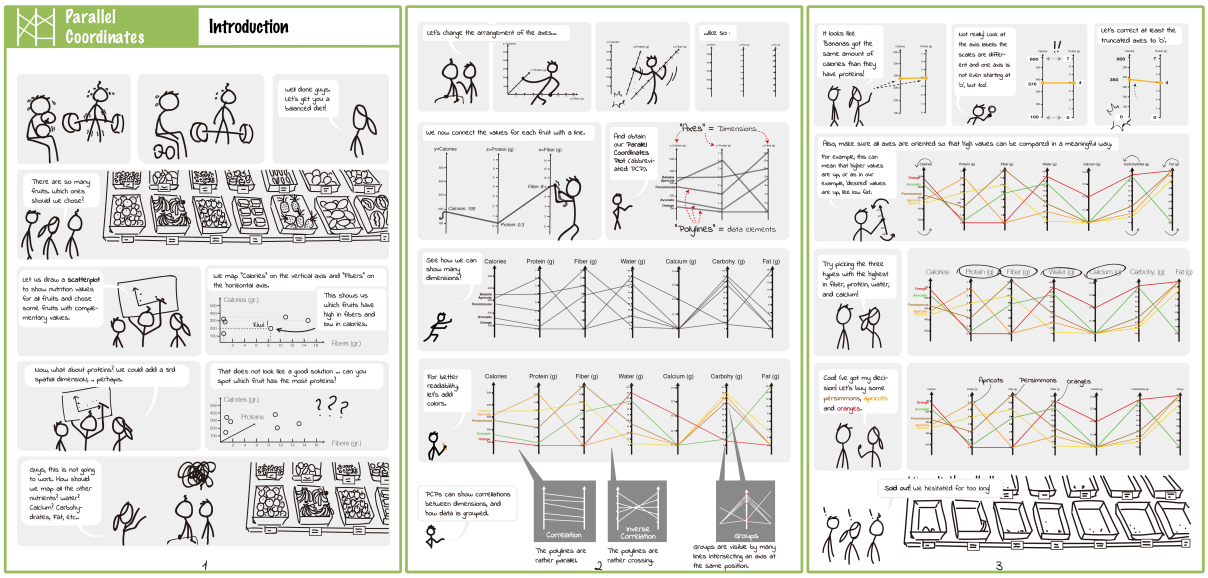


Figure 6.8: *Introduction* for parallel coordinates, featuring the following parts: introducing a data example and giving an example of a common visualization (scatterplot), constructing the new visualization (PCP), explain some patterns and pitfalls, then finish with a joke.

5. **Joke:** Inspired by stories to provoke engagement, a final joke has been highlighted by the participants in our study.

6.7 Evaluation

We conducted a readability study with non-experts, in which participants used our material to support answering questions about three different complex visualisations. This was not a comparative study for two reasons: firstly, as mentioned in Section 6.4, we were carrying out an iterative process involving co-design, so the sessions served to support design iterations (similar to Bach et al. [2016a]) and identify missing patterns or pitfalls, or other issues with the cheat sheets. Secondly, it was not clear which baseline techniques would be appropriate: purely textual techniques would suffer from a lack of the necessary graphical illustrations, while the linear format of video explanation precludes rapid lookup during use.

The study tested for only three visualizations to keep the overall study length to the reasonable time of 1h per participant. We chose the three visualizations with real-world examples for the quiz. During the study, we gave participants cheat sheets for each of three visualization techniques: *boxplots*, parallel coordinates (*PCP*) and adjacency matrices (*matrix*). Participants were given our material and then asked to answer a quiz on each visualization type. The study included two distinct uses of the cheat sheets: *learning*, where the participants read the

Introduction format, and *lookup* where the participants used the material while solving the quiz. During the study, participants were encouraged to annotate the sheets with any omissions, points of confusion or suggestions they had. After the study we gathered subjective feedback on aesthetics, readability, usefulness, and general feedback about preferences and misinterpretations. As highlighted in Section 6.4, we included the feedback from these sessions in our final designs wherever possible.

6.7.1 Participants

Participants were recruited from a local university. To select potential participants with little pre-knowledge about the visualization types that would be tested in the study, we asked respondents to rate their familiarity with the visualizations to be tested on a 5-scale Likert scale (*never seen* (0) to *very familiar* (4)). From the 91 replies we received, we selected the 11 with the lowest rankings. Eight participants rated their familiarity with all visualizations as 0 (*never seen*), while 3 scored 1 for box plots but 0 for the rest. From the 11 selected participants (10 female), 9 were European, 1 Northern American and 1 South African. Four were undergraduates, 3 graduates and 3 doctoral students with background including cognitive science, psychology, philosophy, linguistics art, and university administration staff. Eight participants were native English speakers, 3 staying in a English speaking country for more than half year.

When arriving at the lab, participants were again asked for their familiarity with visualizations in general. Four participants rated themselves 1 out of of 8 on a scale of expertise with 1 being the lowest, 4 rated 2, 1 rated 3 and 1 rated 6. Most participants reported they had only seen graphs like bar and piecharts in news or academic papers. One participant had experience in reading fMRI visualizations but no experience with abstract information visualizations. 5 participants indicated reading visualizations on yearly basis, 1 on weekly basis, 3 on monthly basis and 1 read visualizations daily.

6.7.2 Procedure and Material

For each of the three techniques (*boxplots*, *PCP*, *matrix*) we produced a cheat sheet with a sheet for *Introduction*, *anatomy*, *visual patterns* and *pitfalls*. Each sheet was printed in color on its own A4 sheet. After a brief introduction, participants carried out an exercise with each visualization technique in increasing order of difficulty (based on formal ratings in pilot tests):*boxplot*,*PCP*,*matrix*. For each technique, participants first read the *introduction* while thinking aloud, and were encouraged to annotate anything that was unclear. Next, they were given an example visualization for the respective technique and asked to answer a small quiz

to test their understanding. During the quiz, participants were given printouts of *anatomy*, *visual patterns*, and *pitfalls*. Again, participants were asked to think aloud and annotate the material where unclear. Finally, participants were given a questionnaire asking them to rate *understandability*, *aesthetics*, *usefulness*, *fun*, and *engagingeness* (5-point Likert scale). Each participant obtained £15 for the 1 hour study.

6.7.3 Data and Questions

Each quiz contained between 6 and 7 questions about that particular visualization example, chosen with respect to the cheat sheet components available to the participants. Questions covered low-level reading tasks such as finding single values, making comparisons and identifying patterns. Answers included multiple choice, true-false and short answers, and each question contained an 'I am not sure' option to minimise guessing. The full set of materials and questions can be found in the appendix. In detail:

- We showed 27 **boxplots** of varying shapes, covering the *1908 summer Olympic athlete ages* [Oly, 2019]. Example questions included true or false questions ('More than half of female archery athletes were older than than 40 years old.') and short answer questions ('How many groups were positively skewed?'). Some questions were constructed to be challenging, e.g. 'More than 25% of male boxing athletes were 30 or above 30 years old' required participants to understand overlapping quartile ranges and their meaning.
- for **PCP** we chose two data sets to cover our questions; Fisher's Irish data [fis, 2019] (with 4 dimensions) and the car parameters data set [car, 2019] (8 dimensions). Example questions included 'What's the relation between E and F?' and 'Whats the biggest values recorded at D?'.
- For **matrix**, we chose Les Misèrables co-occurrence network [Bostock, 2012], showing visual patterns such as cliques, clusters, or bridges. Example questions for the *matrix* included 'Are the nodes 14 and 77 connected?' or 'Are cliques A and B connected?'.

As the study went on, we made some modifications to the example datasets to improve readability, as well as better fitting the examples to our tasks. For example, for the *matrix*, we applied an ordering algorithm to facilitate the identification of clusters [Behrisch et al., 2016]. For *boxplot* we removed some overlapping outliers to avoid confusion and allow precise counting where necessary. For *PCP*, we reversed one axis to introduce a typical pitfall. In all cases, we anonymized the data labels.

6.7.4 Results and Findings

Our data included quantitative results from the quiz, subjective feedback, and findings from thinking aloud.

Cheat sheets usability and readability—Subjective feedback on our cheat sheets was very positive. On a scale between -2 and 2, it shows consistently high and very high rankings for understandability, usefulness, and the other measures with rather small confidence intervals (98% Bonferroni-corrected for pairwise comparison). On average, **understandability has been rated 1.6** (Introduction=1.5, *Visual patterns*=1.5, *Anatomy*=1.9, *Pitfalls*=1.6); **aesthetics has been rated 1.8** (*Introduction*=1.8, *Visual patterns*=1.6, *Anatomy*=2.0, *Pitfalls*=1.7); **usefulness has been rated 1.4** (*Introduction*=1.8, *Visual patterns*=1.7, *Anatomy*=1.0, *Pitfalls*=1.1); **fun with 1.8** and **engaging with 1.7** had only been asked for *Introduction*. *Introduction* and *anatomy* received the highest ratings overall: *"Introduction are surprisingly clear. I felt like I understood enough after looking at them."*5. Participants particularly liked narration and humor in the *Introduction* (*"having a story line to follow made it much easier to understand what were the purposes of plotting information in these ways"*5, *"Good, clear illustrations. Humour is important as it is the application to real-life situation."*7. Other comments highlighted the use of *anatomy* and *visual patterns*: *"the anatomy sheets were very useful, the pattern sheets provided the most help when doing the task"*10.

For the individual techniques, those for *matrix* were ranked lower on average, especially the usefulness of *pitfalls* and *anatomy*. From the comments we learned that some concepts ('nodes', 'connections', 'filled-cells') were not explained well on the sheet and that matrices were generally seen as the most complicated visualization (*"The more difficult concept—adjacent matrix was least attractive but I think that is because I find the concept difficult to understand."*7). Despite including specific explanations for it in the *Introduction*, participants had problems to *"determine similarity [between rows] in an adjacency matrix"*10.

Quiz results—Most participants scored well on the quizzes. Correctness-rates were 81% for *boxplot* (distribution of correct answers: **1** **7**), 77% for *PCP* (**1** **7**), and 75% for *matrix* (**1** **6**). For boxplots, the two questions about skewness caused frequent errors: *'More than 25% of male boxing athletes were 30 or above 30 years old.'* and *'How many groups were positively skewed?'* with 6 errors each. Although skewness was explained on the *visual patterns* sheet in both text and picture, from the think-aloud we learned that participants likely ignored the texts. As a consequence, we added larger bold-font keywords and added red color to highlight the message. This leaded to an improved understanding with the last two participants. For *PCP*, the concept of correlation caused problems for some participants (*'What's the relation*

between dimensions B and F?', 'For values increasing at dimension A, do they increase for F as well?'. Since we had only shown patterns for correlation and inverse correlation for neighboring axes, questions that asked for correlations between non-adjacent axes would have required a pattern and explanation we did not consider.

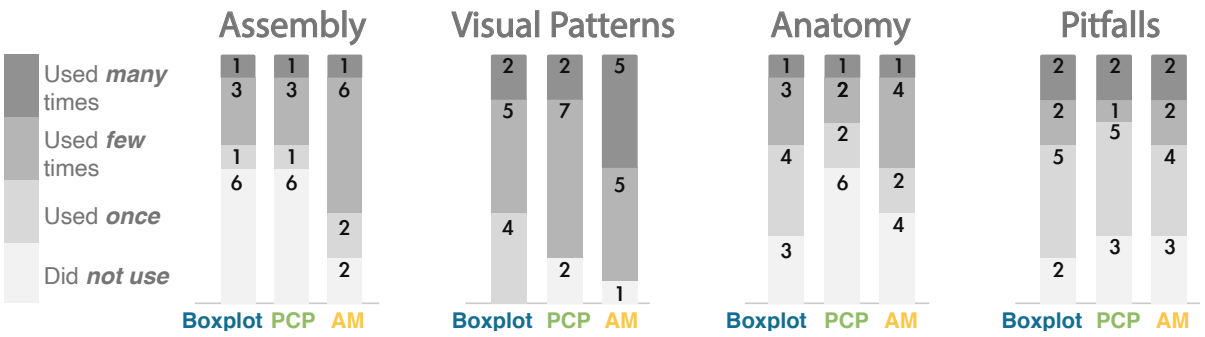


Figure 6.9: Frequencies of using different cheat sheets types for individual techniques during the quiz, as reported by participants.

Cheat sheet use—From observation and think-aloud, we found that all participants used at least one of the modules during the quiz. Usage was highest for *visual patterns* which was used by 90% of the participants, followed by *pitfalls* (75%) and *anatomy* (60%) (Figure 6.9). As expected, *Introduction* was least used (58%) during the quiz since participants already read through the *Introduction* prior to the quiz and had memorized most of the concepts.

Strategies of using the sheets differed somewhat. Before attempting the task, three participants had a glance but did not read the Visualization Cheat Sheets, five read part of the sheets, while three read almost all types of the sheets. When encountering a problem, a majority of participants first tried to solve the questions themselves before consulting the sheets. If they consulted the sheet, they first tried to answer their question from the graphics and only then consulting the textual annotations.

Design changes—We engaged in four major design iterations during the study, informed by participants' feedback. Changes included: highlighting data and terms in *Introduction*; using red to highlight visual elements in pictures with more than two visual hierarchies; adding keywords or annotation to clarify the message; adding terminology for both visual elements and data elements in *anatomy*. We also added variations for each pattern (*"does a cluster have to have self links?"*3), which resolved issues after P3 (*"very easy and quick to find out important information that helped me answer questions"*5.) We found that the 'bridge' pattern on the *visual patterns* of matrix contained too much information at once, so it was split across several panels, despite the additional visual complexity.

6.8 Guideline Workshop

To further evaluate how well our designs and cheatsheets support our goal, we decided to call for another co-design session. This time we prepared a set of design guidelines and visual instructions for creating cheat sheets (call them 'meta cheat-sheets'), to see if these guidelines help participants to create cheat sheets for their own purposes. We created general guidelines as well as guidelines specific to each sheet type with step-by-step instructions, examples, design suggestions, and practices to avoid.

We invited three visualization experts with teaching experience from academia and industry (female, all new to the study), and two participants with more casual experience (non-experts). During the 2-hour workshop, we *i)* introduced our cheat sheets (10min), then *ii)* asked participants to annotate existing cheat sheets for 3 visualisations (30min), then *iii)* discussed usefulness and possible application scenarios, and eventually *iv)* using our guidelines to construct a cheat sheet for a visualisation technique a participant was familiar with.

Similar to the study, participants' annotations and comments showed points of appreciation: *"really clear and explicit description"* (P1) as well as points where additional explanation was needed: *"mention 'Cluster' along an axis after introducing correlation"* (P2) which fed into our design process. Participants appreciated the *introduction* comics, as the stories got past perceptions that *"data is boring and visualization is difficult and I can not do it"* (P1); *"Low in complexity, but easy to understand in here [pointing to her head]"* (P5).

Using our guidelines, participants created *new* cheat sheets from scratch for *timelines*, *connected scatter plots*, *word clouds*, *bar charts* and *sankey diagrams* (see supplementary material). These sheets were coherent with our guidelines; participants organized the contents by dividers or space, and included clear graphics and meaningful annotations. Symbols (arrows, '✓', '✗', '+' and '-') are used in all of their drafts in proper cases. Some participants invented new types of cheat sheets which we will consider to explore in future work, e.g., design variations, colour encodings, and user interactions. Participants also raised open challenges related to using a limited space efficiently for the most important information; preparing materials for other audiences than oneself; and creating *construction* sheets for abstract concepts such as word clouds.

Most importantly, participants reported that the process of creating cheat sheets helped them learn and think about a visualization technique, as the guidelines provided a structure to cover many aspects of a visualization, which evoked actively thinking (e.g., *"if I were to have a test in the next hour, on various types of data visualization, what are the likely test questions?"* (P4). Informed by discussions with the participants, we envisage the following scenarios for using

cheat sheets:

- **Teaching**: visualization cheat sheets provide 'out-of-the-box' teaching materials for presentations in off-line and online courses or workshops, presented and explained on slide shows, handouts or videos (*Anatomy, Construction, Visual Pattern, Well-known relative, Pitfalls*).
- **Learning**: cheat sheets can support independent learning, e.g., the *introduction* introduces basic knowledge about a visualization with comics.
- **Analysis**: Cheat sheets can provide a rapid overview of advanced techniques for analysts to use during exploratory analysis: as print-outs at an analysts desk, or through an online website.

6.9 Discussion and Future Work

Usefulness of Cheat Sheets—Our study set out to show if cheat sheets are a possible way to support visualization literacy, and to refine our final designs and design principles. Our study indicates that we are able to create understandable cheat sheets that novice participants liked and found useful for developing understanding of complex visualisations and that creating cheat sheets helps understanding and memorizing novel techniques. By iterating with experts and novices, we have refined the protocol for creating cheat sheets to enhance clarity and relevance. Together with specific feedback, this has led to the revised guidelines summarised at the end of this section.

Limitations—Our study has two key limitations. Firstly, we cannot say *how much* cheat sheets improved real performance, or which parts were most important. There are many dimensions to any graphical work from aesthetics to conceptual structure, and resource-intensive experimental designs would be needed to truly establish the causal effect of each aspect. Our study did not have a factorial design, but instead explored multiple design decisions at once for the sake of rapid iteration and convergence, for which 11 participants were enough. Moreover, the use of cheat sheets in a lab setting is rather artificial. As a first step, we have investigated cheat sheets in isolation, but there are questions about the complementarities between cheat sheets and other formats (e.g., videos, tutorials) and how these can be brought together for particular learning pathways. However, cheat sheets are intended as concise and structured *support material* for learning, and as a lookup aid when deciphering visualizations, rather than replacements for traditional learning activities and explanations. Future studies should thus assess the use of cheat sheets as general medium for teaching in the classroom or at the office,

and investigate which teaching activities they support best. Eventually, we have used simple, closed tasks, that do not require much context, so future studies should assess how cheat sheets support higher level tasks that require exploration, comprehension and domain knowledge.

Cheat Sheets to Deal with Complexity—The visualization techniques that we are exploring here are relatively complex, both in their presentation and construction. For example, in adjacency matrix, *cliques* are conceptually simple, but the way in which graphics are constructed to make them apparent—re-ordering matrix based on similarity measures, mapping links to cells—is both conceptually and cognitively complex for novices. Participants highlighted these difficulties, but still successfully answered the quiz. Cheat sheets show promise here, as the basic concepts can be presented in a way that helps novices get to grips with the general principles, with more subtle details presented both in patterns and potential pitfalls.

There are many design decisions that have gone into our cheat sheets, and we cannot claim completeness or universality here. In particular, there are many *patterns* that could be highlighted, and their importance is often relative to a particular data context. Similarly, there are decisions to make about whether to classify an example as a *pitfall* or a *variation* (e.g. row ordering in matrices, axis ordering for PCPs). In the absence of categorically correct approaches, we hope that by highlighting both the patterns and the confusions around them, our cheat sheets can help to build up a considered approach to visual interpretation.

Cheat sheet formats—There is potentially a huge design space for cheat sheet formats to explore. Besides *Introduction* and examples in Figure 6.1, examples include posters, slide-shows, videos, or embedded in interactive visualizations as pop-up reading-guides.

Using, Adapting, Improving, Extending—Our cheat sheets are ready to use for everyone and can be found online. They have been designed around simple, extensible templates, with the intention that there is a flexibility in the way in which they are used. The modularity of our cheat sheets makes it possible to assemble their content in new ways (Figure 6.1), collecting and organizing the components to fit a particular audience and format, e.g., posters, mini cheat sheets with only vital information; explanatory notes in infographics or data comics [Bach et al., 2018c]; inserted into slide shows or data videos [Amini et al., 2017].We therefor invite designers, educators, researchers, and practitioners to use our material, to create their own cheat sheets, and to suggest directions for the expansion of cheat sheets that merit exploration. *Firstly*, **more examples of patterns, pitfalls and illustrative examples** will emerge for each sheet. Through experience with deployment and learning, these can be noted and shared. With appropriate support, learners can be enlisted to explain the issues that confused them, adding a broad empirical base to the collection of issues. *Secondly*, **more visualization types** can be added, and we hope that the cheat sheet structure helps the creators of new visualizations (e.g.,

researchers, designers) to bring their work to a wider audience. *Thirdly*, **additional sheet types** can be added, such as explaining interactions or variants of a technique. While we considered some of these in our early designs, we found these types require more research while not being essential to explain a visualization technique and the concept of cheat sheets as in this chapter. More (including non-graphical) information could be added around a visualization technique, such as examples, tools, and data types, explaining algorithms for layouts and re-ordering or covering specific usage strategies such as identifying clusters in matrix or tracing correlation in PCPs. Some of these may be specific to particular visualizations, but practise will help to draw out commonalities. *Finally*, we have worked with a general explanation of each visualization technique, but there is space for additional sheets, or particular configurations aimed at specific contexts such as medical patient data or nutrition. Similarly, high level tasks and specific analysis strategies should be considered.

Towards Guidelines for Creating Cheat Sheets—As already mentioned, there are far too many factors involved in our cheat sheets to assess the value of each individual decision. Importantly, there is a lot of value in *creating* cheat sheets by students as making cheat sheets requires a good understanding of technique, a critical eye, the structuring of knowledge, and a clear presentation. To facilitate the creation of cheat sheets, we summarize some core advice we have learned during our research and the workshops. The full set of instructional material for creating visualization cheat sheets ('meta cheat sheets') including the full list of guidelines and accompanying graphical explanations are found online.

- **Abstraction** is good to make examples general and make this generalizability understood.
- Concrete **examples** help understanding a visualization in the first place such as in our *Introduction*.
- **Visual hierarchy**, using fonts, sizes, gray backgrounds, grouping, and careful use of colors and separators helps structure content and make information retrievable.
- **Maximize graphic-text ratio** to support lookup and visual examplification; use text to give additional information.
- If necessary, **split information** across several comic panels to reduce visual complexity and aid explanation.
- **Show pattern variations** wherever possible to help understanding the main characteristic of a pattern.
- Use **complementary terminology** to help linking concepts from visual space, abstract data space, and concrete application space.

6.10 Conclusion

This chapter introduces the concept of cheat sheets to support literacy around visualization techniques. Each cheat sheet has visual material covering the aspects that are vital for understanding a technique, such as *anatomy*, common *patterns* or *pitfalls*. We defined six types of sheets for six common, yet non-trivial visualization techniques and refined them through an iterative method involving expert interviews and workshops, readability feedback, design-by-example, a readability study, and a guideline workshop. To aid the creation of new cheat sheets, we present a refined set of design guidelines along with our freely available and modifiable library of cheat sheets. We hope that this material helps the community to build a common language and repository of material towards defining and developing a plan for visualization literacy for complex and non-trivial visualization techniques and that this inspires an uptake and popularization of the many different visualization techniques. We also see the potential of using cheat sheeting for inspiring the explanation with visualization in data comics or other formats of narrative with visualizations.

Visualization Cheat Sheets provide another creation reference for data comic creators. Data comics are an effective way to disseminate science in public engagement, with the contribution of this chapter, creators can apply Visualization Cheat Sheets in data comics, e.g., by showing few strips of *construction*; providing *patterns* beside a visualization to aid exploration. Visualization Cheat Sheets can assist audiences with using and understanding data visualization techniques. So, supported by Visualization Cheat Sheets, could data comics be used to present complex data, visualizations, and narratives?

Chapter 7

Data Comics for Reporting Controlled User Studies in Human-Computer Interaction

7.1 Motivation

The topics of content in data comics found in research and in the wild were mainly about climate change, historic events, personal data tracking, and demography. While little practice found that using data comics for scientific reports. We think academic study reporting is potentially benefited by using data comics since academic research reports contain the process of data from collecting, analyzing to result presentation, which is data-driven storytelling in a sense.

Effective communication in academic research is crucial, as it allows readers to assess the rigor of scientific methods and to build trust in scientific results. In particular, controlled experiments require detailed descriptions of key information related to procedures, hypotheses, conditions, data, and experimental setups, as well as careful reporting of data analysis and discussion of the results. The accessible presentation of this information is key not only for peer-review and replication, but for scientific dissemination, the training of students, and communication of research methodology and results to the general public. However, despite guidelines and conventions for technical writing, study reports often result in compressed descriptions that use domain-specific jargon and complex visualizations; important information and methodological decisions risk being buried, complex relations may not be sufficiently explained, or readers might simply overlook information that is crucial for interpretation, discussion, and application of the results [Chalmers and Glasziou, 2009], adding to *research debt* [Olah and Carter, 2017] and creating a need for *interpretive labour* [Graeber, 2015].

To make research findings more accessible to both experts and non-experts, a range of formats have emerged to complement traditional journal and conference articles. Examples include

graphical abstracts [Hullman and Bach, 2018], pictorials, posters, videos, interactive explanations [Victor, 2011a], and academic blogs [res, 2020, med, 2020] which have become one of the key factors initiating and increasing public engagement and effective communication [Kang, 2010, Desrochers et al., 2018]. These formats, often created by professionals [Hullman and Bach, 2018], make compelling use of graphical representations, animation, interaction, and provide more space for detailed and understandable explanations, appropriate for a specific audience. However, this often comes at the cost of over-simplification, which can cause important details about methods, conditions, decisions, and context to be sacrificed in order to create compelling narration and content that is more easily understood.

In seeking to integrate and allow for both detailed presentation of studies and easier access by readers—experts and non-experts—this chapter explores the use of data comics to report controlled user studies in Human-Computer Interaction (HCI) and visualization. Our goal is to provide a framework for researchers and science communicators, guiding the dissemination process and designing visual narratives for scientific study reports (Figure 7.1). We opted for comics because of their great flexibility and numerous characteristics beneficial to communication, including *(i)* tight integration of visual and textual explanations, *(ii)* sequential presentation of information, *(iii)* visual content for recall and quick navigation [Wang et al., 2019c], and *(iv)* different levels of reading, i.e., quick overview as well as access to details, from skimming to close reading. As a static medium, comics are ideal for storyboarding and ideation [Wang et al., 2019b]; there are minimal technical barriers to production, and they can be shared in many forms such as scientific papers, conference posters, slide shows, grant proposals, and blogs. Moreover, comics are different from cartoons, as the basic principles of comics can fit different visual styles to make communication appropriate for a diverse set of audiences and media, e.g., ranging from hand sketches over vector graphics to computer-generated visualizations [Bach et al., 2016a, Kim et al., 2019, Bach et al., 2017-2020]. We focus on exploring the *Visual-Narrative Integration*, *Narrative Structure and Explanatory Visualization* and *Message* elements of the Data Comic Creation Model.

Based on five design objectives we explore the use of comics and data comics design patterns [Bach et al., 2018c] to explain study setups and data analysis for controlled experiments, through a guided design discussion and a series of design iterations focused on previously published studies (Section 7.3). We then elaborate a 10-stage framework to guide the storyboarding of data comics for controlled user studies (Section 7.4), including stages such as *context*, *hypotheses*, *data transformation*, *result presentation*, etc. For each stage, we describe what information should be shown and discuss design solutions that we found in our exploration and working with seven participants in remote collaborative design sessions (Section 7.5). The

collaborative design process and guidelines engaged participants in storyboarding comics for their own studies. We report on visual solutions for common tasks such as sketching hypotheses, reporting randomization, and detailed analysis of results. Feedback from participants highlights the need for improved reporting methods as well as the potential of data comics.

We do not aim to standardize reporting or to provide a visual grammar for using data comics. Nor do we intend to replace other forms of reporting studies and science. Rather, through our framework, which describes critical information that should be retained when simplifying study reporting for visual presentation, we hope to provide guidance and best practices to encourage and inspire researchers and science communicators to engage in both rigorous reporting and efficient explanation of their studies at different levels of detail according to the respective audience. Application scenarios include scientific and public talks, scientific posters, teaching materials, and textbooks. Beyond HCI and visualization, our approach can generalize to other domains, given additional practice and exploration. A comic gallery, workshop material, and stage descriptions can be found at `https://statscomics.github.io`. This chapter contains collaborated works, the author of this thesis is not involved in the creation of Figure 7.1, Figure 7.2, and Figure 7.6.

The original publication of this chapter collaborates with Jacob Ritchie, Jingtao Zhou, Fanny Chevalier, and Benjamin Bach [Wang et al., 2021]. The author of this thesis took major responsibility in the design exploration, created data comic drafts; conducted the collaborative design session; also took part in distilling the design solutions for the common stages, and writing the paper.

7.2 Scientific Reporting

To support writers, reviewers, and readers, numerous international groups have published guidelines to standardize the information that should be provided in publications on the design, conduct, and analysis of the experiments [apa, 2020]. Guidelines exist for clinical trials [Simera et al., 2010], animal research [Kilkenny et al., 2010], observational studies in epidemiology Stroup et al. [2000], and communication of empirical results in HCI [Dragicevic, 2016]—including checklists of information to include when reporting a randomised trial [Group, 2010] or *in-vivo* experiment [Kilkenny et al., 2010]. Some guides prescribe estimation approaches over dichotomous testing procedures for increased interpretability [Dragicevic, 2016]. Other guidelines are established more implicitly through community norms, examples in highly-read papers, and good scholarly training.

Still, technical reports may not always result in a clear and comprehensive account of the

research [O'Donnell, 2005]. Moreover, as publication migrates online, it has become increasingly easy to append supplemental materials to research articles. Open data repositories, project websites, and data analysis scripts are now commonly used as a transparent demonstration of materials and study procedures to promote replication [Hornbæk et al., 2014]. While this is extremely valuable, it adds to the complexity of information, and given the increasing pace of publication in many fields, even trained scientists can experience problems with understanding reported information and interpreting the output of statistical tools. Misconceptions and fallacies [Dragicevic, 2016] as well as an increasing demand for the piecing together of information from multiple sources can lead to misguided or incorrect conclusions [Kline, 2004, Beyth-Marom et al., 2008], and well-documented problems such as *publication bias* and the *file drawer effect* inflate the number of incorrect claims in the literature [Begg, 1994, Rothstein et al., 2005, Scargle, 1999].

In our work, we go beyond the traditional article format, exploring ways to report experiments—in line with scientific standards—that promote understandable communication of scientific knowledge while preserving transparency.

7.2.1 Presentation and Explanation

In addition to prose, graphical content is key to reporting scientific methods and results [Lynch and Woolgar, 1990, Pauwels, 2006]. Figures in the form of graphs, diagrams, pictures, illustrations, and other data visualizations show experimental procedures [Karaca, 2017], study setups [Huron et al., 2016], theoretical models [Barberousse, 2013], and research hypotheses [Ho et al., 2017, Endicott et al., 2010]. Besides diagrams and illustrations, data visualizations are a key means of conveying results graphically by showing mean values, distributions, confidence intervals, outliers and other sources of uncertainty. Symbolic representations, e.g., for confidence intervals or quartiles in boxplots, can provide hurdles for novice readers [Correll and Gleicher, 2014] and should be explained [Wang et al., 2019b].

There have been a number of suggestions to change or augment scientific reporting for better explanation. For example, Document Cards [Strobelt et al., 2009] provide for automatic summaries of scientific papers, including keywords and figures to facilitate browsing large document collections for relevance. However, our focus is not on concise summaries but on understandable explanations. In this vein, fluid documents [Zellweger et al., 1998, 2002] allow readers to pull up supplemental information in-context, whereas elastic documents [Badam et al., 2019] support linking of textual and tabular content to automatically generated visualizations. Other new formats harness the ability of papers in HTML and PDF to include animation and

interactivity to convey dynamic behaviours in user interfaces [Grossman et al., 2015] or let readers explore variability in result outcomes across multiple analysis settings [Dragicevic et al., 2019]. Similarly, the need to understand complicated concepts in machine learning and data science paired with an increasing (public, scientific, juristic, etc.) demand for transparency has led to an increase in interactive *explorable explanations*. Pioneered by Victor [2011a,b] and now widely disseminated by platforms such as `Distill.pub` [Dis, 2020], explorable explanations leverage interactive visualizations to explain complex concepts [See, 2020]. In general, while there exist many other sources to explain scientific methods (t-tests, ANOVA, etc.) our work focuses on the analysis pipeline—its stages and the decisions made for the conditions, factors, and setup, as well as the reporting of results—rather than explaining the functioning of a specific aspect of, e.g., a statistical model.

Our work is closer to graphical abstracts, a particular and increasingly common format to illustrate the core idea of a research paper through a single and concise graphic that can shows concepts, research methods, experimental setups, and results [Hullman and Bach, 2018]. Comics, in comparison, provide for more space and narrative elements to deal with complex information and to enable a more in-depth understanding of a study's protocol and results.

The approach we present in this chapter differs from such uses of comics to teach statistics or visualization, since the focus is not on educating or explaining particular statistical methods, but on reporting how such methods are *used* in a specific scientific analysis, to allow a reader to understand the purpose they serve. This is closer to the concept of a *data comic* [Zhao et al., 2015b, Bach et al., 2017a], a genre that focuses on effective narration and explanation with data visualizations. Data comics offer a large design space comprising many data comic design patterns [Bach et al., 2018c] and explored in previous workshops [Wang et al., 2019b].

Thus, rather than understanding comics as sketchy 'cartoons' or as just an educational resource for public outreach, we explore comics as an effective means of communicating key information and decisions for reporting controlled, quantitative empirical studies in human computer interaction (HCI) and visualization. There is a low barrier to creating publication-ready comics, as well as universal shareability through paper publications, conference posters, textbooks, slideshows and websites. To the best of our knowledge, no examples, structured design explorations, or studies exist on the topic of using comics to report on scientific studies.

7.3 Data Comics for reporting studies

7.3.1 Design Objectives

In exploring data comics for controlled user studies in HCI and visualization, we aim to investigate challenges and design solutions for the following *design objectives O1-O5*. Each objective marks an individual aspect of our approach to addressing study reporting.

O1: Clarify decisions and important information: We aim to clarify key steps and decisions across methods, protocols, and results. This includes information about study setup, number and demographics of participants, conditions and datasets used in trials, user interfaces, or data transformations such as outlier removal. Each of these decisions has implications for interpreting the results. Our goal is to present and provide for visual explanation (O2) of these decisions to facilitate the interpretation of results.

O2: Explain information visually: Our exploration focuses on effective visual explanation of the information in a study report. Where appropriate, we want to find visual depictions to explain hypotheses, data samples, study protocols, task conditions and data transformations and to explain results through data visualization. The repertoire of visual depictions may range from diagrams and schemata, to graphic pictograms and annotations, screenshots, illustrations, and data visualizations. Textual or other narrative structures present in comics are outside our scope.

O3: Support recognition and recall with a visual vocabulary: Besides using visual explanations (O2) for understanding key information (O1) and providing for overview (O4), visual information can afford quick reference through recall and recognition. For example, an effect discussed in the results section of a study report may be present only for a specific condition, which can be reviewed by glancing back to the relevant section of the comic. A visual vocabulary made up of recognizable symbols for each condition might facilitate this process. While earlier work has found benefits of data comics for memorizing content [Wang et al., 2019c], investigating memorability of comics is beyond this chapter's scope.

O4: Provide visual overview and structure to facilitate information retrieval: For example, individual stages of a study or analysis could inform the glanceable higher level structure of the comic. At the same time, we aim to provide space for sufficiently detailed, specific information, e.g., though dedicated panels of different sizes. Reading the comic at different levels of detail could help make comics-style study reports compelling to different audiences, including both experts and non-experts.

O5: Use aesthetic appeal to create engagement: Finally, besides the rather functional objectives O1-O4, we posit that reporting studies through comics can provide for a fresh view of science through compelling visual narratives and clear visual styles (e.g., hand-drawn illustrations and characters). We believe comics—provided the right visual style for different contents and audience groups—can engage non-experts and larger audiences, further promoting science outreach.

7.3.2 Methodology

Our methodology was driven by the need to create comics for studies, due to the lack of existing examples. Following a method used in prior research Bach et al. [2016a, 2018c], we began by creating comics for studies reported in existing academic papers and eventually created 15 comic sketches for 13 papers, 4 of which we polished and which can be found on our website. Our detailed methodology was as follows.

1. **Design Exploration:** We started our exploration by creating comics of our own papers that involved empirical studies, since we were most familiar with the intricate details of these experiments: [Bach et al., 2018a, Wang et al., 2019c, Chevalier et al., 2014, Alper et al., 2013, Ritchie et al., 2019]. As our research lies within visualization and HCI, our papers inherently contained a number of helpful visual resources that were used as the basis of the visual vocabulary. We constructed at least one complete example comic for each study, starting with sketching on paper, and later translating to sketching on drawing boards, which allowed for quick iteration and exploration of ideas, layouts, and visual styles. During this iterative process which involved discussions among all authors, we were aided by data comic design patterns [Bach et al., 2018c] and existing guidelines for visually explaining data visualization techniques [Wang et al., 2020].
2. **Design Discussion:** In parallel to our own exploration, we conducted one 1.5 hour design discussion with three participants (two graduate students specializing in HCI and one undergraduate), who were asked to create data comics for a research paper that they were familiar with, with an objective of using visual explanations to support understanding of the analysis. The discussion was introduced using a presentation, during which participants were shown examples of data comics, facilitated by one of the authors. Participants first created written lists of all steps of the analysis, which they then turned into comic sketches with pen and paper.

   Resulting data comics showed integrated use of visual and textual explanation, with a clear reading order indicated by comic panels (details are discussed in section 7.6). These comics

provide visual solutions for illustrating data collection, using annotated illustrations (e.g., interaction between a user and a smart watch), task procedures described using a simple flowchart, and interpretations of study results (e.g., a bar chart showing an overall trend, with annotated details beside one of the bars). All three participants included narrator characters to explain study details. In the discussion following the workshop, participants commented that data comics could reveal concrete information such as the effects of outlier removal or enable critical reasoning about whether 'outliers' were really outliers. However, participants reported that they struggled to decide which parts of the analysis should be included into the comic. One reason mentioned was the lack of familiarity with studies: *"my level of comfort with stats is too low to do the translation to comic manually"*. Such comments motivated us to identify information to be included before conducting any further evaluation (section 7.4).

3. **Design by stages and examples:** To promote transparency and the reporting of key information, we consulted existing checklists and guidelines for scientific reporting (section 7.2). We analysed examples produced during our exploration and design discussion, and distilled an initial list of stages and design solutions reported in section 7.4. To evaluate whether our stages could effectively serve as a prescriptive guide to creating comic-style study reports, we asked a computer science student with design skills who was unaware of our research to create more data comics. The generated examples clearly presented the structure of the stages and effectively applied design solutions.

   To further explore the usefulness of data comics in presenting research, we extended our practice to HCI studies about music [Simon et al., 2008] and speech language processing [Ruan et al., 2018]. The inclusion of a particular paper was based on two principles: *i)* it included a controlled user study, and *ii)* we were familiar with the topic so that we could maximize our effectiveness in explaining the hypotheses, procedures and analyses. During this process we kept developing and refining our list of stages and associated design solutions (section 7.4).
4. **Collaborative design:** Eventually, to further evaluate our stages and the idea of using data comics for reporting, to source more design solutions, as well as to find problems and/or solutions not yet covered by our own explorations, we conducted another set of collaborative design sessions with seven researchers (five doctoral and two post-graduate) (Section 7.5).

7.3.3 Example Comic: Weighted Graph Comparison

To better explain what information our comics aim to encompass, and how this information is visually presented with respect to our design objectives O1-O5 (subsection 7.3.1), this subsection discusses an example data comic for a published study. Figure 7.1 shows a comic illustrating a

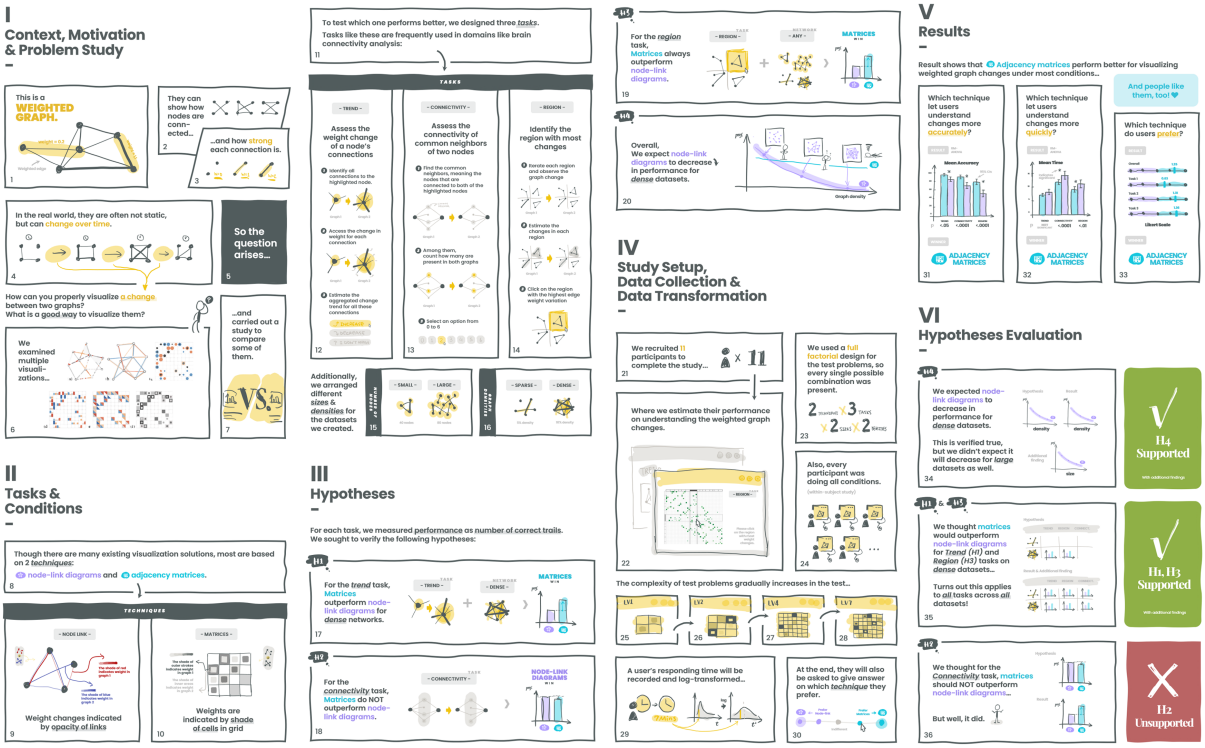


Figure 7.1: Example of a data comic illustrating a study that compared two visualization techniques for weighted networks. Created by one of the coauthors through an iterative process, based on information provided in the original paper [Alper et al., 2013]. Each panel is labeled with a number in the bottom left corner, which we reference in (subsection 7.3.3).

study on comparing weighted networks as conducted by Alper et al. [2013]. The paper presents a classic example of a controlled user study in visualization; in a lab setting, it compares two visualization techniques and tests four hypotheses on task accuracy and task-completion time. The comic is meant for a blog, project website, or conference poster.

Structure: The comic in Figure 7.1 is subdivided into four parts `I-IV` by the author to highlight important stages of the study report (O1). Panel numbers 1-36 are defined in the bottom left corner for each panel of the comic. The comic begins (`I`) by introducing **context and study motivation**, including key concepts (e.g., weighted graphs in panels 1-4), purpose of the study and research question 4. It also introduces the subject of study, i.e., the six visualizations of interest 6. Both concepts and conditions (I) are introduced with illustrations which are reused throughout the remainder of the comic, providing visual consistency and quick visual reference. Then (`II`), the comic introduces **tasks and conditions** examined in the study. It does so by introducing formal notions for the two conditions and visual explanations for each condition 9-10, using the *anatomy* cheat sheet described by Wang et al. [2020]. Each task 12-14 is explained through a vertical sequence of images explaining the task's goal, what participants

had to look for, how they entered their answers, and the set of possible answers. The last panel in part II introduces the different data sets and which parameters were used to generate or select them.

Part III of this comic describes each of the four **hypotheses** 17-20 for the study by repeating visual depictions of relevant conditions involved in each hypothesis (task, datasets). In addition, each hypothesis illustrates the expected results in terms of the dependent variable, if that hypothesis should hold, using annotated bar chart sketches. Part IV details the **study setup**, including participants, setup, conditions, and a reconstructed screenshot of the interface presented to participants 21-22. Panels 25-28 show that the order in which data sets were shown goes from simple to difficult. The last two panels in IV illustrate that task completion time has been log-transformed 29 and that participants were asked for a subjective ranking on a Likert scale 30.

The results section of the comic (V) displays **results** for each of the independent variables, highlights significant findings, and "declares" the winner technique for independent variables. This part uses a parallel layout 31-33 to repeat and stress that *"adjacency matrices outperform node-link diagrams"*, a conclusion drawn from the experiment and also emphasized as a corresponding core finding in the original publication. Finally (VI), the comic **evaluates the hypotheses** and reports which ones are supported and which are unsupported.

Design decisions: The comic makes a set of design decisions to communicate study design and results. As a form of sequential art, comics are naturally averse to organizing hierarchies, so extra care has been taken to ensure clear section structures and semantic nestings. The heading for each part, for example, used large spacing and typography to preserve proper contrast in the masonry layout, aiming for visual clarity and structure (O4). To demarcate the canvas regions and bounds of tasks and conditions, the author uses characteristic panels to group each of them 9-10, 12-14, 15, and 16. The primary condition "techniques," i.e., node-link diagrams and adjacency matrices, are given different identity colors to be reused 8 (O3). A sketch of the user interface in the experiment is placed in the study setup section to give the audience a clearer understanding of the tasks 22 (O2). The comic also employs a set of data comic design patterns [Bach et al., 2018c], such as *flashback* 4, *space-annotation* 20, *temporal sequence* 6, 25-28, *question-and-answer* and *multiple facets* 31-33.

The visual style (O5) of the comic aims to achieve a balance between clarity (sufficient spacing, reduced and harmonic color palette, consistent visual emphasis, clean, grid-like, and non-overlapping layout, abstraction of concepts), objectivity (few and very small narrators), as well as artistic demand to create a compelling and unique visual experience (hand-drawn style, pictograms).

	Stage	Explains	Design Solutions and Examples
1	**Context**	Problem, motivation, jargons & concepts, domain, background	Explaining domain jargon, sketch context (Figure. 7.2)
2	**Conditions**	Techniques, devices	Condition Symbols (Figure. 7.3(b) Figure. 7.3(c))
3	**Hypotheses**	Hypotheses, expected results	Sketch Hypothesis (Figure. 7.4(b) Figure. 7.4(a))
4	**Tasks**	Tasks, instructions, answer modalities, collected data	Illustrate Task (Figure. 7.5(a))
5	**Stimuli and Materials**	Data sets, user interface	Explain the conditions in Stimuli (Figure. 7.5(b))
6	**Participants**	Power analysis, eligibility, demographics, numbers	Randomization Plot (Figure. 7.6)
7	**Study Setup**	Study environment, steps (training, tasks), study design	Flowchart or timeline (Figure. 7.7)
8	**Data Transformations**	Assumption checks (e.g., normality), transformations (e.g., z-scoring, logarithmic), outlier removal	Show Transformation (Figure. 7.8(b) Figure. 7.8(c)) Zoom (Figure. 7.8(c))
9	**Result Presentation**	Explain important messages of a chart, explain error bars	Difference heatmap (Figure. 7.9(b), Tips for interpreting the charts Figure. 7.9(a))
10	**Hypotheses Evaluation**	Which hypotheses are supported by results	Comparing the sketch from hypotheses with the chart from result (Figure. 7.1 34-36)

Table 7.1: Overview of the 10 framework stages (column with a head of **Stage**), including what information is reported (column headed **Explains**) and the **Design Solutions and Examples** we found.

7.4 Stages and Design Solutions

This section describes the 10 stages which informed the design of our data comics study reports. In our framework, a stage summarizes the information a data comic can show about a specific aspect of the reporting process, and provides visual solutions for common communication needs. A stage can be seen as a short set of guidelines about *what* information to report in each part of a study and *how* to represent this information through graphics, text, and a combination thereof. We grouped information into stages based on common reporting structures in HCI and visualization papers, our experience about what information can effectively be shown in context (e.g., tasks, independent variables), and existing guidelines (Section 7.2).

Each stage is illustrated with small comic strips selected from example comics created by the authors of this chapter and works co-designed with participants (Section 7.5). The selected examples demonstrate *possible* design solutions, and we extract commonly used abstract design patterns from the examples if they exist. For each stage we provide a brief definition, explain the presentation objective and discuss challenges and design considerations. Table 7.1 summarizes the information presented at each stage and our visual solutions.

Stage #1—Context and Motivation

Purpose: This stage aims at a general introduction to the problem studied, the motivation for the study and the research question, and explains any important jargon occurring in the remainder of the study. Depending on the intended audience and their knowledge, additional information about the topic and domain might be necessary. **Design:** This stage also introduces the visual morphology and terminology for the entire comic, e.g., by explaining concepts visually. For example, Figure 7.2 and Figure 7.1[1]-[4] both introduce basic terms (i.e., 'transition' between 'initial visual state' and 'final visual state'; 'weighted graph') through visual explanations using sequences, arrows, and textual annotations. Keywords are highlighted with color to visually link to corresponding visual explanations, such as 'transition' and the '→' in Figure 7.2 and 'a change' and '→' in Figure 7.1. The key problems a study addresses can be explicitly stated, as in Figure 7.2.

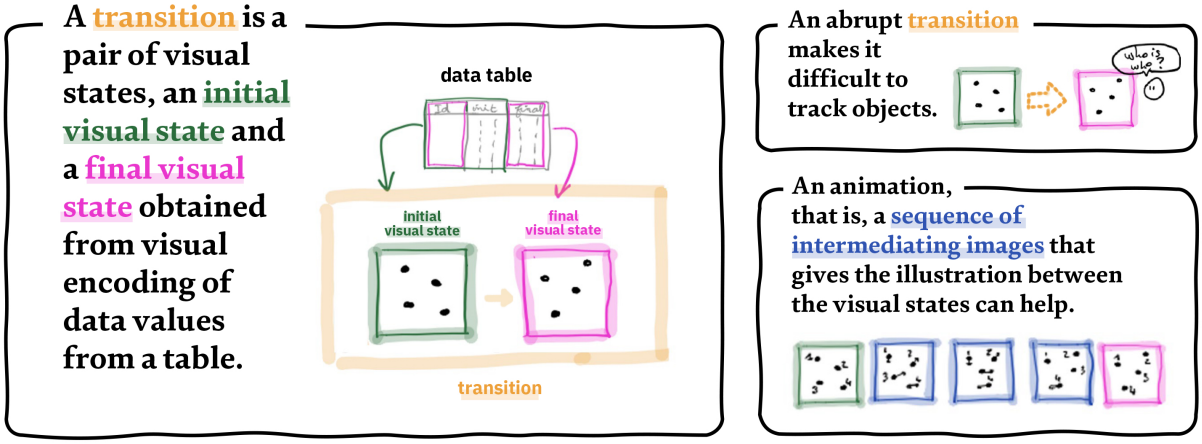


Figure 7.2: Example for Stage 1: The comic highlights and explains key concepts ('transition', 'initial visual state', 'final visual state', 'animation') [Chevalier et al., 2014] with illustrations of simple examples.

Stage #2—Conditions

Purpose: This stage introduces the conditions in a study, i.e., the values chosen for independent study variables. In HCI and visualization this usually includes comparing visualization and interaction techniques, devices, or subsets of a design space. This stage should explain how each condition looks and works, and which ones are the baseline techniques, explaining and illustrating each technique. **Design:** Conditions can be illustrated through schematic representations (Figures 7.3(a), 7.3(b), 7.3(c)) or, if more iconicity is required, screenshots, photos

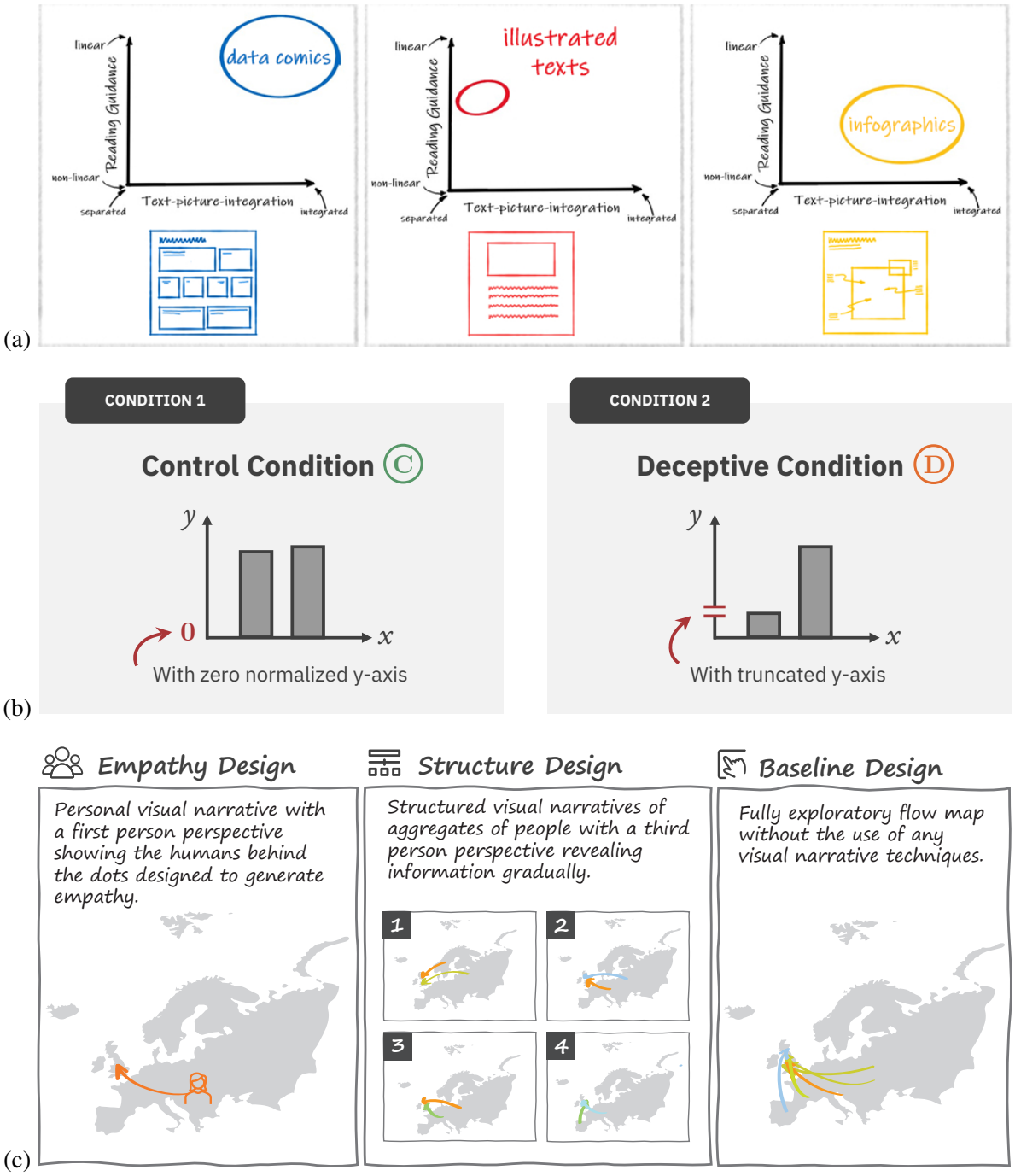


Figure 7.3: Examples for Stage 2: Descriptions and iconic condition symbols for (a) visualization types [Wang et al., 2019c] (b) visual encodings [Ritchie et al., 2019] (c) two visual narratives compared to a baseline [Liem et al., 2020].

and other realistic representations. Techniques should be explained well, and for visualization techniques the author should explain how to read each visualization, e.g., through anatomy cheat sheets [Wang et al., 2020]. Figure 7.3(b) shows two tested conditions side-by-side, highlighting

the differences in visual encoding (truncated vs. non-truncated y-axis). Figure 7.3(a) shows how different techniques relate to a (Cartesian) design space. Symbols can be defined here and re-used to indicate each condition for the remainder of the comic, e.g., using '©' to present 'control condition' and '' to present 'Empathy Design'.

Stage #3—Hypotheses

Purpose: This stage illustrates and describes the hypotheses tested in the study. A hypothesis can involve one or more conditions, at least one dependent variable, as well as expected outcomes stated in terms of effect size. **Design:** Illustrating hypotheses can be challenging due to their *hypothetical* nature. Illustrating the conditions involved can be straightforward, e.g., by repeating figures introduced for conditions (and potentially also for tasks, Figure 7.4(b) and Figure 7.4(a)). To illustrate hypothetical values for dependent variables, we can use data visualizations with the same encoding as those used to report results, as this makes evaluating hypotheses easier. For a single quantitative dependent variable, we tried bar charts. If a study author pre-registers a hypothesis and a minimum expected effect size, this is straightforward to visualize. However, this is not a well-adopted research practice, and less well-defined hypotheses such as 'larger than', 'no difference' or 'range' are fuzzy and hard to convey in a single bar chart. Annotations could highlight concepts such as differences, averages, values and associated ranges, such as Figure 7.4(a) which uses '' to indicate 'less'. Visualizing such concepts is related to visualizing uncertainty, for which visualization techniques have been summarized elsewhere [Hullman et al., 2019]. One solution is to omit axis tick marks along axes (Figures 7.4(b) and 7.4(a)). Similar presentation is found in research papers [Ho et al., 2017], for example, Endicott et al. [2010] uses schematic diagrams to illustrate hypotheses about the origin of the modern human.

Stage #4—Tasks and Dependent Variables

Purpose: This stage illustrates the individual tasks participants perform during the study. This can include instructions given to participants, the way instructions were presented, what participants saw in each task, what they had to do, how they entered an answer to a question and completed a trial, which answers were valid, what data were collected for each dependent variable and how these data were collected. **Design:** Tasks can be shown through a sequence of actions (Figure 7.5(b)), by showing example stimuli, what users needed to pay attention to or count (Figure 7.1[12]-[14]) and eventually showing the interaction used to finish a trial and submit an answer (e.g., Figure 7.5(b) shows a clicking action, while Figure 7.1 (bottom of [12]-[14]) shows a multiple choice menu).

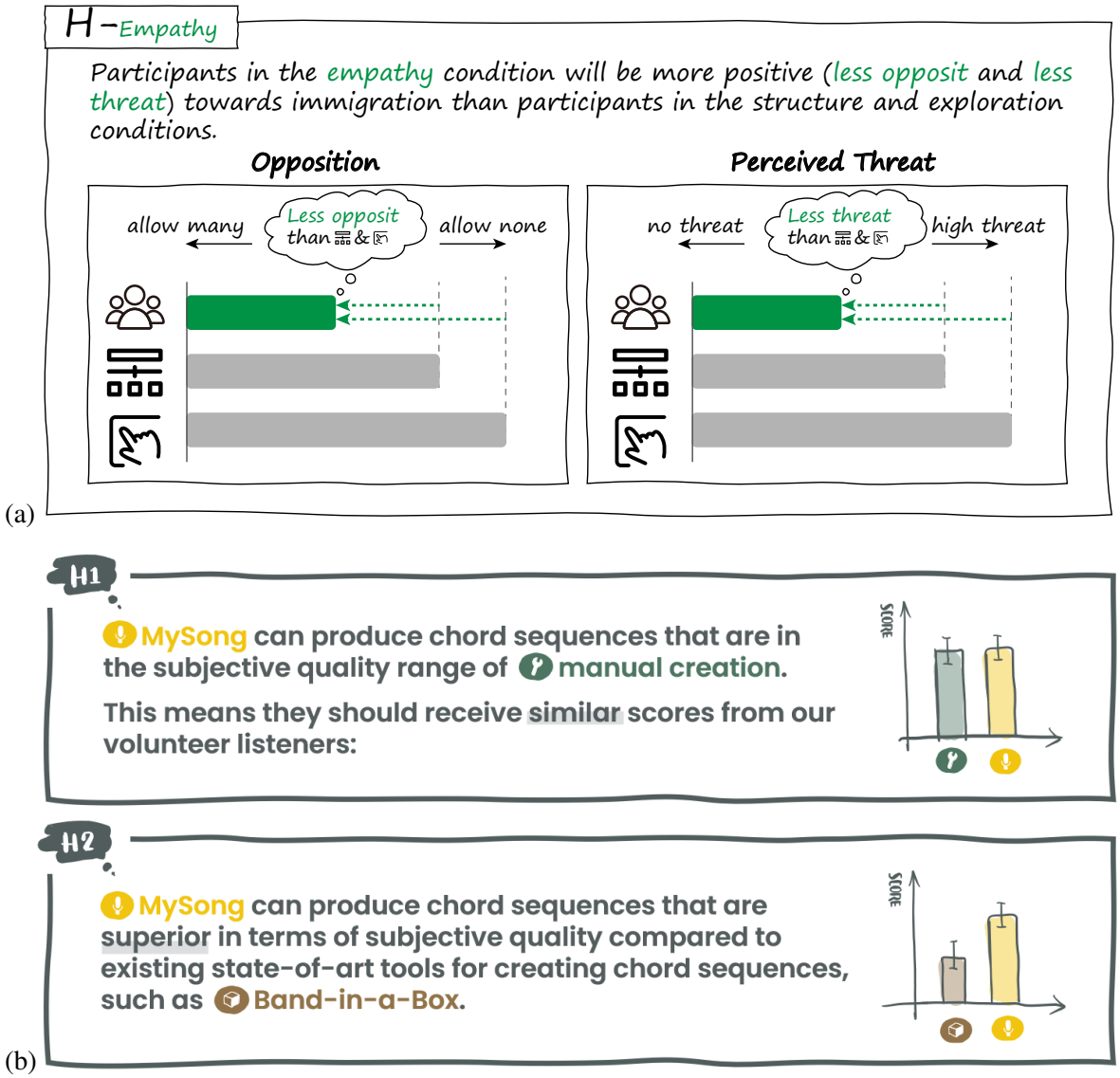


Figure 7.4: Examples for Stage 3: Descriptions and sketches of hypotheses for (a) visualization study [Liem et al., 2020], (b) musical interface evaluation [Simon et al., 2008].

Stage #5—Stimuli and Materials

Purpose: This stage should introduce the individual stimuli presented to participants during the study. This is related to the previous stage on *Tasks and Dependent Variables* but it aims to give examples of stimuli and explains the factors involved in their design. For example, in visualization, this should include the example data sets in the setup participants saw and details like their size and complexity. If data sets are generated, illustrations could show parameters and values. **Design:** Factors used to generate data sets can be shown by examples, such as Figure 7.1[15]-[16] which shows examples of small and large, dense and sparse networks or

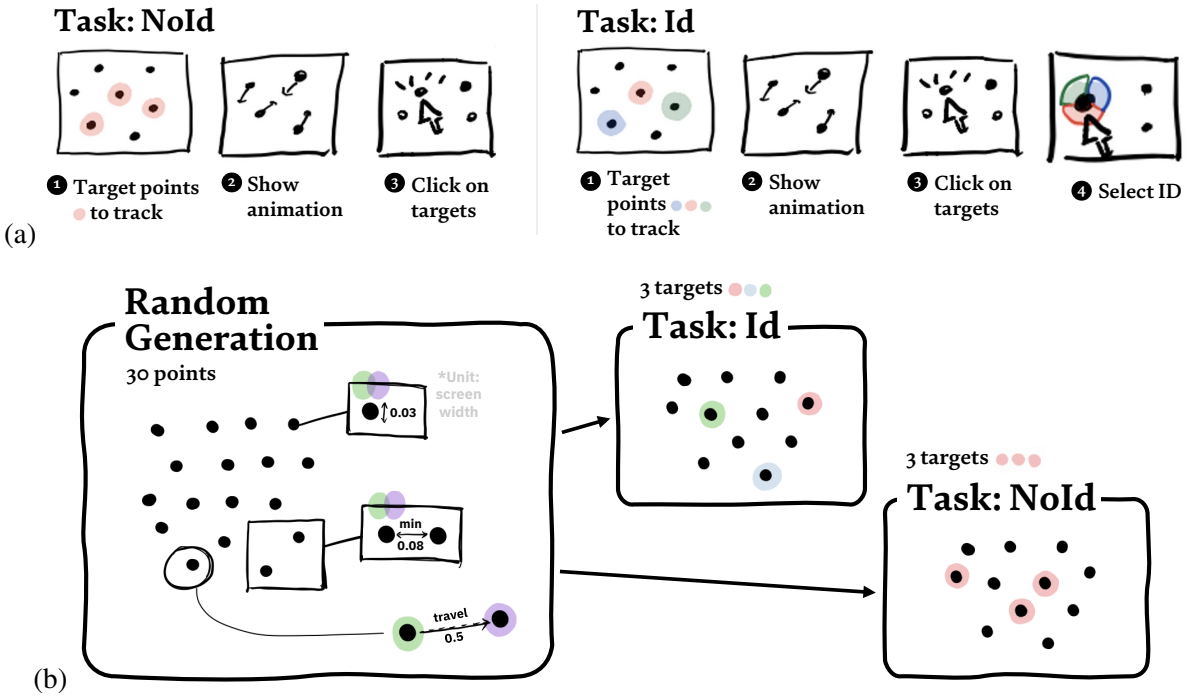


Figure 7.5: Example for Stage 4 and Stage 5: (a) how participants in a perceptual study [Chevalier et al., 2014] select targets in two conditions (b) how the stimuli are generated for these tasks.

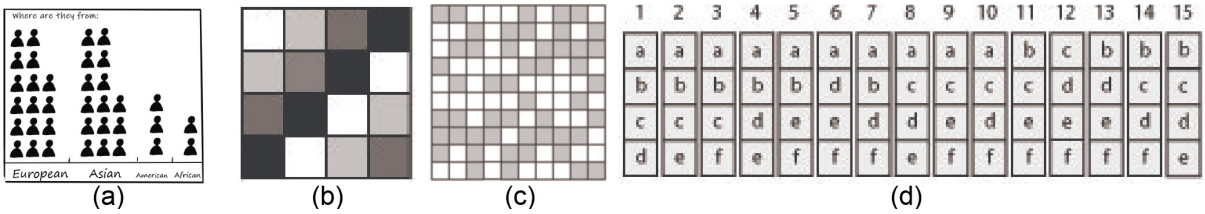


Figure 7.6: Examples for Stage 6: (a) participant demographics isotype chart [Wang et al., 2019c], (b) Latin Square randomization, (c) completely randomized design and (d) balanced incomplete block design from [Yarnes, 2013].

Figure 5 of Bach et al. [2018a] showing example stimuli with different parameters (number-of-clusters) as well as the variation of data sets with the same parameters ((c), (d)). The example in Figure 7.5(b) shows an example of a data set with randomly generated point positions. Multiple small panels are used to display detailed views of different conditions.

Stage #6—Participants

Purpose: This stage should include information about participants themselves (e.g., number, gender balance, eligibility criteria), as well as results of any power analysis used to estimate the number of participants required to detect an effect. **Design:** Figure 7.6(a) shows participant

demographic data using isotype visualization. A randomization plot can be employed to show various randomization schemes (Figure. 7.6 (b) Latin Square randomization [Yarnes, 2013] (c) completely randomized design [Exp, 2020] or (d) balanced incomplete block design [Yarnes, 2013]) to demonstrate the allocation of participants to stimuli. If power analysis is used to determine sample size, a power curve for the study design [Eiselmayer et al., 2019] can be shown to display the power estimate and its margin of error.

Stage #7—Study Setup

Purpose: This stage helps the reader get an overview of the study procedure. This can include the study environment and apparatus, the sequence of steps in a study (e.g., training, trials) as well as detailed information associated with each step (e.g., duration, setup) and the instructions participants were provided with at each step. **Design:** The sequence of steps can be shown as a pipeline (Figure 7.7(b)) or a flowchart (Figure 7.7(a)). Annotations provide both an overview of the pipeline and detailed information about participants, materials and duration. Study setup can be complex, and may be explained over several sections in a paper, so the author may be selective about the information reported in this stage.

Stage #8—Data Transformations and Checks

Purpose: This stage describes transformations (e.g., log-transform) and checks performed on the data *before* any significance tests or regression analyses. This stage can include outlier removal and winsorizing, or checking for normality before running statistical tests. **Design:** Outlier removal can be indicated by thresholds, e.g., at the extreme of a distribution as in Figure 7.8(b), and placed alongside descriptions of the nature of the outliers (e.g., as explanations in speech bubbles). Figure 7.8(c) shows a distribution before and after transformation; possible visualizations include histograms or $Q-Q$ plots.

Stage #9—Result Presentation

Purpose: Compared to the previous stages, this stage may require far greater consideration, given the importance of figures and visualizations for analysis and presentation of results. The information presented can include anything important for understanding the results: distributions, effect sizes, descriptive statistics, error bars or measures of uncertainty. Data visualizations are key in this stage and require appropriate descriptions as to how they should be read [Wang et al., 2020] as well as annotations to highlight key messages. **Design:** A number of different

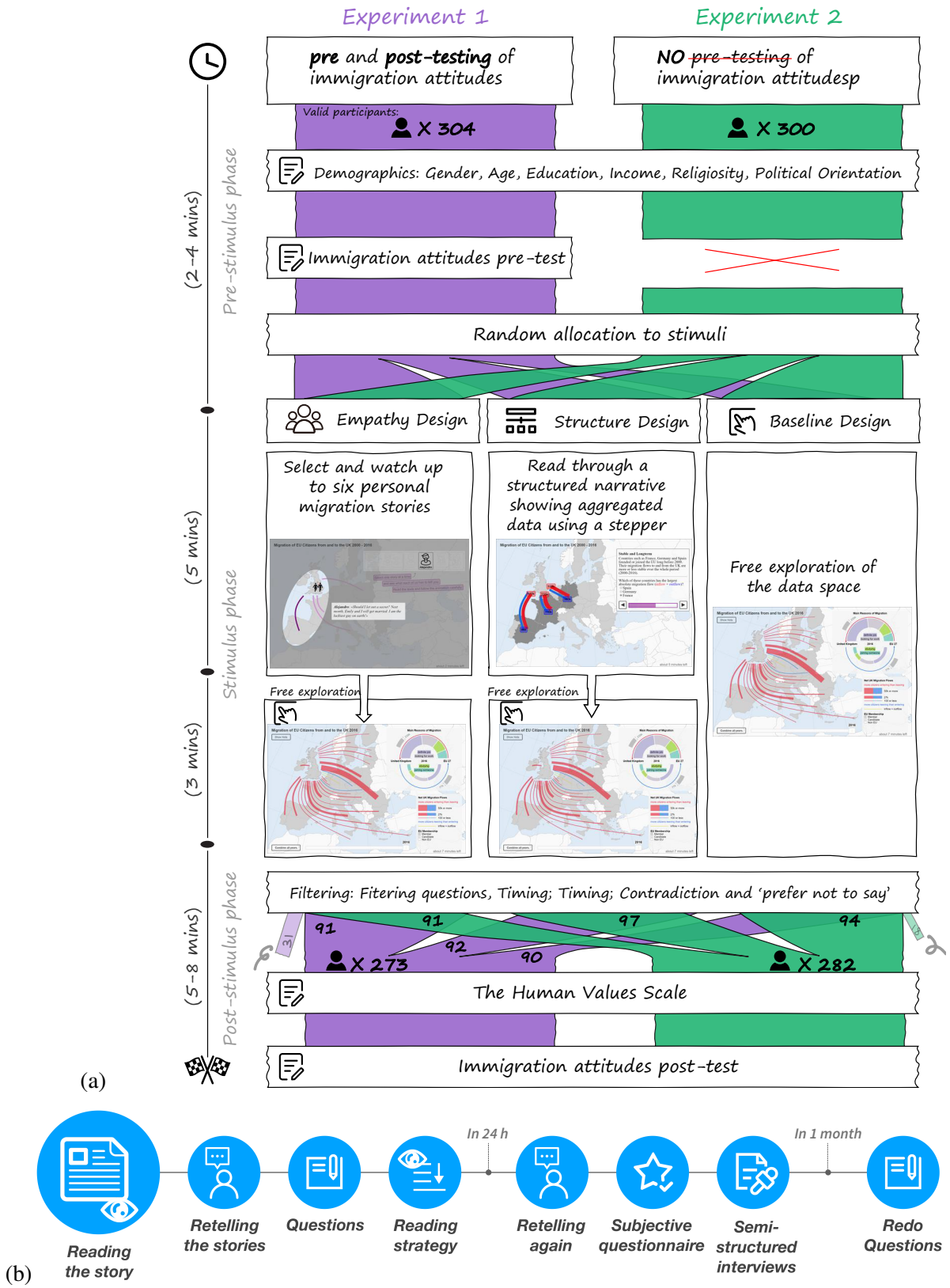


Figure 7.7: Examples for Stage 7: (a) pipeline describes the sequence of study steps and how procedures differed between conditions across two experiments [Liem et al., 2020] (b) A flowchart describes the different phases of a study in chronological order [Wang et al., 2019c].

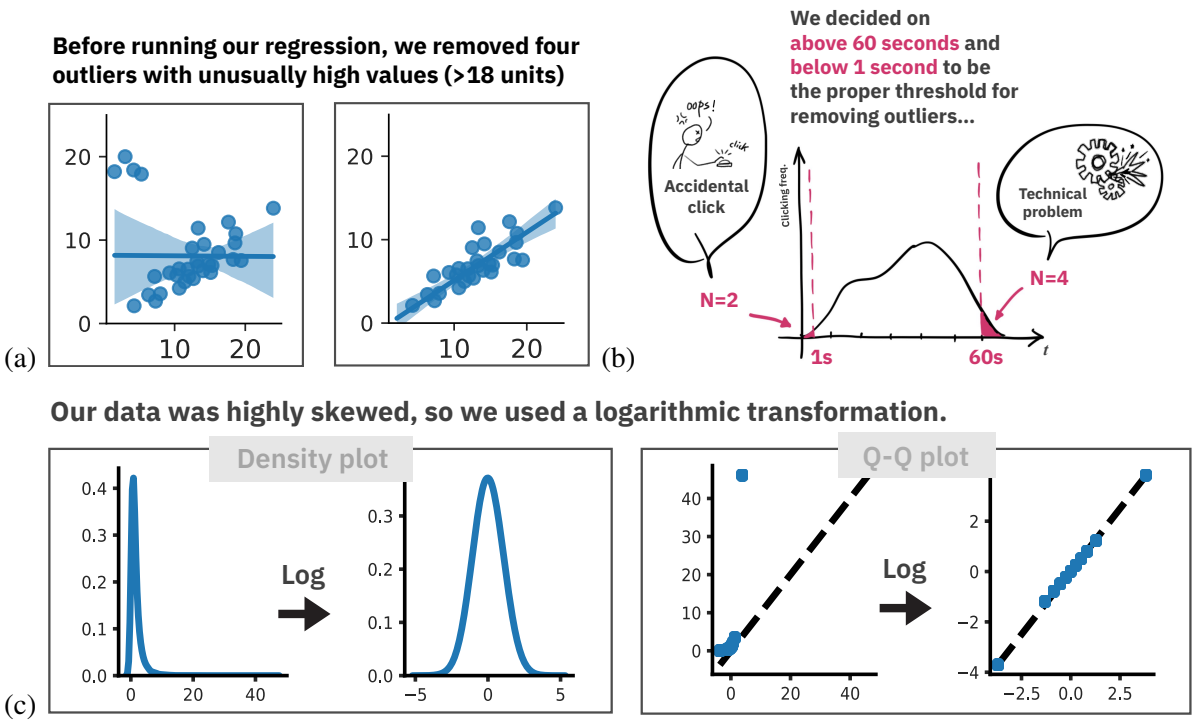


Figure 7.8: Examples for Stage 8: (a) Visualizing the process of outlier removal reveals that the goodness of fit in a regression model was highly dependent on outlier removal. (b) Authors can justify criteria for outlier removal using annotated density plots [Bach et al., 2018a] (c) Data comics can demonstrate the effect of important distribution checks and transformations that are often relegated to a textual description.

techniques may be necessary for this stage: for example, bars in a result bar chart can be reordered to facilitate pair-wise comparison of distributions. Such a reordering can be made explicit through a before-after sequence and, e.g., an arrow highlighting the change (Figure 7.9(d)(d)). Figure 7.9(c)(c) uses a cut-out panel to reveal a bimodal distribution with a violin plot. Other statistical results might require showing trends, extreme values, distributions or outliers, which can be facilitated by annotations. In case of complex charts or insights, multiple repetitions of a chart can be presented, each one focusing on a specific message (Figure. 7.9(b)). To show pair-wise differences between more than two independent variables, Figure. 7.9(a) uses a "difference heat-map", showing pair-wise differences on different levels (.01, .001).

Stage #10—Hypothesis Evaluation

Purpose: This last stage contrasts the initial hypotheses with the findings from the results in #9. It should show which hypotheses can be accepted and which ones are not supported by the

Figure 7.9: Examples for Stage 9: (a). A complex result [Ritchie et al., 2019] can be presented using a sequence of secondary visualizations (b) or by using multiple explanations of a repeated chart to focus on specific messages. (c) A Zoom pattern [Bach et al., 2018c] can be used to reveal distribution details (d) and panel sequences can illustrate operations like variable re-ordering.

results. **Design:** Figure. 7.1, for example, shows both, the figure drawn for the initial hypotheses (#2) and the the actual result figure. Supported and unsupported hypotheses can be highlighted with '✓' and '✗'.

7.5 Collaborative Design Sessions

We conducted remote collaborative design sessions to evaluate our solutions and stages and obtain feedback, explore more solutions, and to understand which challenges people might encounter in creating data comics for a wider range of studies.

7.5.1 Participants

Participants were recruited through email lists at computer science research labs at three universities (two in Europe and one in Canada). We motivated participation by providing guidance on how to create data comics for the students' own studies, co-designed and polished by two of the authors. We gave the comic in Figure 7.1 as an example. We did not constrain the scope to controlled studies in HCI and visualization. The seven participants (three females) include five Ph.D. students and two recently graduated master students specialising in usable security and privacy (*SecPriv*), computational social psychology (*CompPsy*), geoinformatics (*GeoInf*), probabilistic machine learning (*ML*), 3D interactive visualization (*3DVis*), public health and health system (*Health*), and systems design engineering (*SysEng*) respectively. Their respective studies involved a controlled study in data visualization (*GeoInf*,[Liem et al., 2020]), an analysis of online open human data. (*SecPriv*[Tahaei et al., 2020]; *CompPsy*,[Chen et al., 2020]), observational animal data and modeling (*ML*, [Camilleri and Williams, 2020]), 3D interactive system design and evaluation (*3DVis*, work-in-progress), cognitive behavior study with a game (*Health*, [Chan, Long Ting, 2019]) and a study of gestural interactions with VR environments (*SysEng*, work-in-progress). This selection gave us the chance to explore the potential and limits of our design solutions and receive feedback from diverse perspectives. None of the participants had any experience with creating (data) comics.

7.5.2 Procedure

The entire design session was conducted individually for each participant, giving participants and facilitators a flexible schedule to work on comics asynchronously offline and communicate about solutions and design iterations.

1	*What motivated you in participating in this workshop?*
2	*Which issues do you currently see in reporting studies (as a reader/author)?*
3	*Did our stages-framework help? Which challenges did you experience in doing your story boards / what was the most difficult part, e.g., stages, illustrating specific issues?*
4	*Creating storyboards, did that change the way you think about reporting studies?*
5	*How are you going to use your final comic?*
6	*What did you value about using data comics to reporting studies, what do you think works about using comics for reporting studies?*

Table 7.2: Questions asked of the participants in the co-design process.

1. For the first session, participants were asked to attend with a write-up of their study and to give a brief presentation in 5-10 minutes. Three participants presented slides. This gave facilitators and participants an overview and a review of the important stages of the study.
2. Then, we gave a very brief introduction to data comics and showed examples from `https://statscomics.github.io/gallery.html`.
3. Participants were then asked to create drafts ("storybords") for their comics off-line before the next session. Participants were free to choose drawing techniques, e.g., hand-sketches, vector graphics, photoshop, screenshots, etc). To guide participants and to see if our framework would help them in their process, we provided textual guidelines of each stages (`https://statscomics.github.io/guidelines.html`) as well as the three example comics.
4. After 1-4 days, instructors met again with each participant, discussing storyboards and any challenges encountered.
5. Based on feedback from participants, the instructors iterated upon their comics, aided by slides, written reports and other records.
6. Eventually, participants were asked summative questions (Table 2).

7.5.3 Results

Drafts from participants showed that they presented a visual and structured layout of their studies. The structure included our stages, *e.g., motivation*, *methods*, *conditions*, *stimulus* and *results*. Design solutions included *a)* a Venn diagram to represent the relationship between data and samples (*SecPriv*, *Health*); *b)* a timeline to illustrate tasks (*GeoInf*, *CompPsy*, *3DVis*); *c)* simple pictographs to demonstrate differences between technical terms (*GeoInf*, *CompPsy*, *Health*, *3DVis*, *SysEng*) (e.g.,'mood' and 'emotion'), and *d)* illustration and interpretation of the conditions and materials used in the study (*GeoInf*, *SysEng*). All participants presented

annotated result charts interpreting features such as significant differences (*GeoInf*), error bars (*ML*) and the presence of *'less fluctuation'* and *'more fluctuation'* in a Gaussian process regression (*CompPsy*).

7.5.4 Qualitative feedback

All participants in the collaborative design session expressed very positive comments about the experience and the results of the session, describing data comics as, e.g., a "visual interface into a more written format" (*GeoInf*). Participants were **motivated to participate** in the collaborative design session as it was an *"interesting problem"* to *"expand [the] visibility of [their] research"* and to enable *"more people get the message"* (*CompPsy*). Participants found data comics a *"very good opportunity to [create] a visual, easy to consume format for a blog post"* (*GeoInf*).

About **current ways of reporting**, participants found the complexity of the information a challenge: *"it is often very time consuming to collect all the valuable/required information about a study reported in a classical paper format [...] one has to 'close read' the paper and investigate additional material to find this information, and [expand] the effort to find the right information."* (*GeoInf*). From an author's perspective, the drawbacks of current reporting include that it is difficult to *"convey/communicate the complexity of study designs and/or stimuli in a compelling, easy to understand format."* (*GeoInf*). Moreover, there is little support for grasping the information in a paper: *"the current expected paper format/structure does not encourage authors to provide a (standardised) high-level yet detailed summary of such information in a 'front-matter' of a paper (besides the abstract, where detail is limited; or teaser-images)."* (*GeoInf*).

For the **eventual use of their comic**, participants agreed on general uses of *"outreach, publicity"* (*GeoInf*). More concrete scenarios included the use of comics as a *"script for [a] conference presentation"* (*ML*) and as suitable material for blog posts, posters (*"instead of posters I can also reuse data comics."* (*SecPriv*)), conference presentations and online project pages to *"present what we did[,] in a 5 minute-read"* (*SecPriv*). For future, wider use, one participant wished that data comics *"would become an accepted format [...] for publishing research"* and mentioned potential extensions through *"interactive formats like scroll[ing] stories"* (*GeoInf*). This participant also acknowledged the need for guidance and standardization *"similar to pre-registration templates or the stages-framework used in the [co-design workshop]."*

With respect to the **instructional material** provided from our side, participants found our motivating examples *"well-illustrated"* and highlighted that the feedback from instructors was very helpful *"since this challenges my perception of other people's understanding"* (*CompPsy*).

Our stages helped participants *"[c]oming up with an idea of what I want"* (*SecPriv*), to *"distill the required information"* and create *"a structure to follow"* (*GeoInf*). However two participants found it easy to follow the stages since they had already written up their study reports. Stages were also found to be helpful to guide the inclusion or exlusion of information, since they helped participants to focus on the process of textual and visual explanation, as well as *"to think about all the other possibilities comics have to communicate a narrative."* (*GeoInf*).

Creating comics involved a set of **challenges**. Despite positive feedback on our stages, three participants reported issues with deciding what information should be included and identifying the target audience to get *"the message across"* (*CompPsy*). This is a common issue in storytelling, which goes beyond the specific domain of visualization. Drawing with pen and paper was reported as an issue by one participant. With respect to individual stages, we found most difficulties arose with stage #5 (*Stimuli and Materials*, *"explain[ing] multiple datasets/methods"*), stage #7 (*Study Setup*, *"complexity of study designs and the complexity of stimuli"*) (*GeoInf*) and stage #8 (*Data Transformations and Checks*,*"explaining transformations/re-binning of collected information"*). Although the stages were given as guidance, one participant found difficulties in deciding *"what sequence was the best, the amount of text to include and [the way of] making it easy-to-understand"* (*SysEng*). A specific challenge arose from the need to explain specific machine learning models (*ML*), algorithms (*CompPsy*) and formulas (*SysEng*). While outside the scope of this chapter, this will require significant attention in future work.

Asked whether **creating storyboards changed participants' thinking** about reporting studies, the answers indicated that participants found *"storytelling is important"* (*Health*) and data comics a *"very suitable format"* (*GeoInf*) "*on top of my list of methods/techniques to report findings or study methods.*" (*SysEng*) and would suggest such comics *"as a service for researchers"* (*SecPriv*) since they *"capture the idea of my study in a very short period of time and the idea stays in their mind"* (*ML*).

7.6 Discussion and Future Work

7.6.1 Designing Comics for Studies:

Our collaborative design sessions showed that researchers unfamiliar with general visualization practices clearly see the need for better communication and the potential of techniques such as data comics. Moreover, our stages discussed in section 7.4, alongside examples, helped participants understand this novel medium and guide them in expressing their own ideas visually as storyboards. As crafting sophisticated visual content was found to be an intimidating task

by some participants, our collaborative sessions employed a 2-step iterative co-design process, comprising of clear instructions and guidelines about what was expected from a participant (stages, storyboard) and the involvement of a professional designer polishing the storyboard and helping with design solutions. We believe this process can be refined for future, including more extensive workshops for illustrating studies and for data comics in general [Wang et al., 2019b]. Further workshops and teaching practice will help refine our methodologies and build a formal set of guidelines, instructions, and design templates. An open question is to what extent we can guide decisions on level-of-detail and better design data comics for specific audiences.

7.6.2 Design Objectives:

Our collection of comics represents a series of points in the solution space for visually explaining procedures and findings in comic-style study reports. We use this section to reflect on how these relate to our initial design objectives (**O1-O5**).

Our first design goals were aimed at finding ways to visually explain (**O2**) key concepts (**O1**). These concepts were represented through different types of graphics; for example, iconic illustrations were used to display contextual messages related to the goals of the study. Schematics were used to present study methodologies or task workflows. Screenshots were used to present realistic representations of prototype UIs and interactions performed by study participants. Where iconic representations where not possible, e.g., due to the visual space required (e.g., graph sizes in Figure 7.1[15]-[16]) or the abstract nature of the concept (e.g., a technique) the designer used more symbolic representations to create a visual vocabulary (**O3**). Data visualizations were naturally used to present results, but also to show demographics and data transformations. Some figures explain visualizations and charts, e.g., error-bars, chart axes, or user interface components (*GeoInf*).

We found some solutions were able to serve as patterns or templates, similar to data comics design patterns [Bach et al., 2018c], supporting transparency and understanding in visual explanations. Examples include step-by-step explanations of tasks (e,g, Figure 7.5(a)) making evidence for key claims visually apparent, e.g., facilitating comparison by reordering the bars in a bar chart (Figure 7.9(d)); and evaluating hypotheses by comparing schematic visualizations of hypotheses with the actual results (Figure 7.1).

Our criteria of visual overview (**O4**) refers to the ability to provide information at different levels of detail through overview and details. For example, by structuring a comic layout into larger and smaller panels, headlines and subsections, and nested subpanels, and placing panels in clear sequence (Figure 7.1[25]-[28]) to show a temporal relation, or in parallel (Figure 7.1[31]-[33])

to show information of parallel meaning. For complex layouts, arrows indicate reading direction and avoid "backlocks" [Cohn, 2013a].

We seek visual consistency in illustration style; colors used for highlighting and indicating different conditions; and visual lexicons for individual objects or actions in a comic (**O3**). This visual vocabulary, which is gradually introduced to the audience in the early stages of the narrative framework, allows readers to build associations between different experimental concepts and to quickly refer back.

It should be noted that using conventional comic-style presentation does not indicate poor quality (**O5**). This aesthetic is chosen with the goal of complementing the textual annotations, and reducing visuals that are irrelevant to a viewer's understanding of the study report. While the style of data comics was not focus of our work, we believe there are many open research questions about the impact of the visual style for understanding, as well as on attractiveness and credibility [Wood et al., 2012].

7.6.3 Limitations and Open Questions:

The research presented in this chapter is clearly limited by the **number of examples** we could produce and the number of interactions we could have with the participants in our design discussion and collaborative design sessions. Unlike for infographics [Chevalier et al., 2013, Borkin et al., 2013, Lu et al., 2020], data videos [Amini et al., 2015], or data comics [Bach et al., 2017-2020], there is no corpus that we could curate and learn from to understand best practices, or the effectiveness of existing data comics for reporting empirical studies. Our research has created an initial set of such examples, developed through design exploration and iterative collaborative design, that we make available on our website: `https://statscomics.github.io`. Our results are only the beginning of a much larger exploration of this vast design space that is yet to be structured and formally evaluated through, *e.g.,* controlled user studies to assess which of our solutions work best and to investigate the effect of various factors on effective scientific communication for various audiences with respect to comprehension, memorability, engagement, and educational value.

Providing **better support for authoring** data comics is another item to be added to the research agenda. Creating comics involves many skills [Wang et al., 2019b] and authoring support for specialized comics such as data comics [Kim et al., 2019] is still in its infancy. Possibly, authoring tools could be integrated with data exploration and analysis platforms such as Python notebooks. One challenge with creating such tools is to find the right balance between expressive power and usability, as well as between guidance and creative freedom [Shneiderman,

2007]. This is further complicated by the wide range of elements and concepts in study reports, and the rich visual vocabulary afforded by data comics.

More detailed and specialized guidelines for common study protocols (e.g., Fitt's Law studies) could also help address **challenges** raised by the participants, some of whom found complex study designs difficult to communicate. Guidelines could also be refined according to a particular comic's intended use case, by indicating to authors the appropriate level of simplification for their audience. Without additional guidance, there is the risk that authors might leave out important details when attempting to communicate to a non-technical audience.

We have deliberately left out in our work the **explanation of statistical and other tests, algorithms, and complex systems** (though this was mentioned as a difficult point in the process by several participants). For example, while descriptive statistics and regression analysis are commonly explained using graphics, more sophisticated methods pose challenges to understanding and transparency, and could benefit from research into how they could be communicated using comics.

7.6.4 Towards Better Support for Reporting Studies.

During our research, we engaged in co-design with researchers from economics, cyber-security, computational social psychology, machine learning, geo-informatics, 3D scientific visualization, and HCI. In this chapter we provide design solutions of what one *could* do. While some fields rely on data analysis and modelling uncommon in HCI, others conduct mostly qualitative research, where data comprises of interview records and other observations. We do not yet provide specific visual solutions for such scenarios, but these examples can inform future research.

Moreover, the boundaries between comics for reporting studies and engaging in public outreach begin to blur, e.g., PhD Comics [Cham] and xkcd [Munroe] are examples of science-related comics that often "go viral" [Cham, 2012]. We believe that, as comics are a universal communication medium deeply rooted in human culture, data comics (and their close cousins, infographics and graphical abstracts) will receive greater consideration in the near future, helping to narrow the gaps between scientists and the general public, and between communicators and audiences. These formats can inform the design of better graphics in papers, appendices, web pages, talks and slideshows, posters, and videos. Actively engaging in making visual information more commonplace can be transformative to communication: as scientists begin to think about their research more visually, the media for communicating and transmitting thoughts will evolve, and the audience will in turn become more visually literate. We encourage researchers

and science communicators to seek inspiration in our work to go beyond the preconceptions associated with comics: data comics and sketching are less about cartoons or particular visual styles and more about thinking, logic, and effective communication. We hope our work will contribute to a new culture of reporting and discussing scientific findings, and to a discussion of transparency, teaching statistics, and data literacy in general.

7.7 Conclusion

This chapter explores data comics for illustrating reports of controlled studies. Through an iterative co-design process, we defined common design solutions, guided by five design objectives. Our results and co-design experience are encouraging and can inform further research on improving scientific reporting. We contribute a set of examples of how comics support that reporting, providing an initial set of concrete ideas for how to communicate common stages of HCI studies and statistical analysis, and may serve as a starting point for future exploration. Finally, we show that our new method allows people to create clear and compelling visual summaries of complex research designs.

Data comics for reporting controlled user studies are our attempt of applying data comics to solve the real-world communication challenge, combining the knowledge learned from previous chapters, we take a close look at data comic design solutions for specific types of messages (i.e. the stages of Section 7.4). The examples presented in this chapter may be used as design templates for future data comic reproduction. However, as the complexity of a story increases, a data comic can take larger space than the traditional text-based reports. We see a need for methods that present information in a glanceable and selectable way so that the audience can quickly navigate to the information they need, for example, to switch between different data sets and analysis techniques; jump to a certain stage of a report; to test if keeping or removing an outlier can influence the result. Making data comics interactive can be a solution to enable exploration and more active reading.

Chapter 8

Interactive Data Comics

8.1 Motivation

So far, we have explored several essential elements from the Data Comic Creation Model, including reading sequence and levels of picture-and-text integration (*Visual-Narrative Integration*) in chapter 4; created data comics with participants in a workshop setting (Chapter 5); provided creation supports include data comic design patterns [Bach et al., 2018c] for panel transition and layout (*Visual-Narrative Integration*), data visualization cheat sheets for interpreting visualization techniques (*Narrative Structure and Explanatory Visualization*) in chapter 6; applied data comics to report the common stages of controlled user-study (*Visual-Narrative Integration*, *Narrative Structure and Explanatory Visualization* and *Message*).

However, data comics we have created, encountered in the wild [Cagle, 2019, Bach et al., 2017-2020] and research are mostly static; the panel layout itself encodes the flow of information, guiding the reader through a predefined sequence of panels [Manning, 1998]. A major advantage of static comics compared to other storytelling mediums is the absence of programming or animation skills required for authoring them, making them more approachable to a wide audience of creators.

Conversely, interactivity is a core element in understanding data through visualization and enables viewers to consume a story in different ways, giving more agency around how the narrative unfolds, and eventually may improve comprehension and recall in storytelling [Kim et al., 2017c, Romat et al., 2020, McKenna et al., 2017a]. Interactivity in storytelling enables the viewer to e.g., drill into parts of a story to access additional information, context, or explanations Chang et al. [2000], Badam et al. [2019], Victor [2011a], or to curate their own version of the story given their own approach to exploring the data [Dragicevic et al., 2019]. The comic artist community has already experimented with interactivity in creative ways such as

letting the reader navigate an infinite canvas [McCloud, 2000], or linking between nested panels to create multiple narrative paths for a personalized story [Goodbrey, 2013]. These compelling explorations hint at a larger design space for adding interaction to aid storytelling with data comics.

In this chapter, we contribute the formalization of a core set of operations for interactive data comics, which can help inform the development of future graphical user interfaces for creation tools. We propose a lightweight scripting approach enabling to add these operations to a set of existing comic panels. Triggered by a user interaction, an operation transforms one or more of the comic components: content, structure, or visual appearance. These operations dynamically append and replace panels, load and switch comic layouts, highlight and filter elements, as well as enable data input and manipulation.

To inform the set and design of interactive operations, we establish six goals of interactivity for data-driven storytelling, and report on a systematic review of traditional (non-data) comics to understand which interactions the community of comics authors is currently providing to support some of these goals (section 8.4). When attempting to design and create interactive comics building on these operations, we found no appropriate tool to support rapid prototyping for story authors (section 8.5). Creating interactive and potentially non-linear data comics in prototyping tools such as Adobe XD or Figma leads to exponentially complex views and a substantial performance challenge. The only existing alternatives are full-fledged programming approaches (e.g., HTML, CSS, and JavaScript), which are less approachable to many designers through their steep learning curve, but also require developers significant development time.

To mitigate this problem and allow story authors to more quickly explore interactivity, we propose a lightweight declarative scripting approach: COMICSCRIPT (section 8.6). COMICSCRIPT is grounded in our formalization and operationalization of the core building blocks for interactive data comics. COMICSCRIPT is inspired by other declarative languages such as Vega-Lite [Satyanarayan et al., 2017] and DataTheater [Lau and Guo, 2020]. In our approach, designers first create panels in traditional graphics editing tools or digitize hand-drawn sketches. Then, they define comic layouts and *script* interaction in a web editor. Thus, rather than specifying graphical elements, COMICSCRIPT allows to augment existing graphics with interactivity using a core set of operations.

To explore the design space of interactive operations and iterate on COMICSCRIPT, we engaged in a two-week-long design process with six illustrators, designers, and postgraduate computer science students (section 8.7). Feedback suggests that our scripting approach is understandable by non-programmers and fosters exploration of the potential of interaction for data comics. Our study shows potential for interactive storytelling, while providing an extensible

basis for exploring future authoring tools and types of interaction. It suggests possibilities for a range of scenarios such as exploration, personalization, learning data literacy as well as engaging viewers with messages. A gallery of interactive data comics can be found online, alongside the editor, tutorials, and detailed documentation: `https://interactivedatacomics.github.io`.

The original publication of this chapter collaborates with Hugo Romat, Fanny Chevalier, Nathalie Henry Riche, Dave Murray-Rust, and Benjamin Bach [Wang et al., 2022]. The author of this thesis took the main responsibility for designing the interactions in data comics, making interactive prototypes, distilling the operations, and conducting the workshop; took part in preparing the workshop, and writing the paper. The author of this thesis is not involved in the coding process presented in section 8.4, and programming of the online editor and COMICSCRIPT, while the function of the editor and COMICSCRIPT is informed by the interactive operation and examples created by the author of this thesis.

8.2 Interaction in Storytelling

Interaction plays a crucial role in information visualization, enabling analysts and domain scientists to explore multiple facets of their data and discover patterns [Card, 1999, Heer and Shneiderman, 2012]. Interactive visualization is also valuable to a more general audience and a variety of tasks and contexts [Elmqvist et al., 2011] such as searching histories of names [Wattenberg, 2005], browsing through a corpus of books [Thudt et al., 2011], or learning experiences in museums [Hinrichs et al., 2008]. Interaction is prevalent in data-driven storytelling [Riche et al., 2018], whether it allows the viewer to enhance story structure and enable reader-driven navigation, link different story elements [Sultanum et al., 2021], or provide controlled explorations such as dynamic queries [Segel and Heer, 2010, Stolper et al., 2018]. In fact, most data-driven stories encountered in the wild from news media outlets, designers and practitioners provide some degree of interaction (see, e.g., [Hohman et al., 2020] for a comprehensive list).

The value of interaction in storytelling is manyfold: it is used to engage viewers more deeply with the data to enhance their comprehension and memorability, to prompting self-reflection, foster critical thinking, or personalize reading, to name a few [Hohman et al., 2020, McKenna et al., 2017a]. Depending on the type of message evidenced by the data, viewers may learn something new [Yee and Chu, 2015, Hilton et al., 2020], change their perspective on a topic [The New York Times, 2015, 2017], or modify their behavior [The Washington Post, 2020]. A few research studies demonstrated several of the direct benefits of adding interaction to data-driven stories. Kim et al. [2017c] demonstrated that interactions eliciting the viewer to externalize their

prior knowledge before contrasting it with the data enabled a deeper comprehension and recall of the information. Romat et al. [2020] found that providing interactions to personalize visual elements was likely to impact how viewers related to data and encouraged deeper reflection. Zhi et al. [2019] demonstrated that interactions may also impact levels of user engagement.

As pointed by Amanda Cox in her keynote at VisWeek 2011 [Cox, 2011], one of the challenges of interactive storytelling is how viewers discover the interactive elements and how to use them. Boy et al. [2016] provide solutions to enhance affordances and make these interactions discoverable. Other research shows that different interaction mechanisms may appeal to different people, making it difficult to recommend specific types of interactions [McKenna et al., 2017a].

To summarize, while there are challenges to creating appropriate, compelling interactions for storytelling media and making them discoverable, several studies demonstrate direct benefits to readers. Interactive experiences may connect readers more deeply to the data, enhance comprehension and recall of data-driven stories and open up new kinds of engagement. This present work seeks to explore what interactions can be added to the data comics medium, and how it could be done.

Data comics are typically created as 'static' pieces, which makes the creation and sharing easy compared to dynamic work such as animations and interactive media. Comics can be read at an individual pace: a reader can pause, reflect, glance back to previous panels, or compare panels. Large and information-rich panels can invite a reader into exploration and discovery. Small multiples, as used in exploratory visualization [Bach et al., 2015, Lekschas et al., 2021], can provide the reader with an *at-a-glance* overview of actors, data, visualization, insights, and events in a data story, facilitating individual access. While authors such as Chris Ware have been experimenting with non-linear storytelling in static comics, adding dynamic interaction to paper versions allows more possibilities [Andrews et al., 2012].

In this chapter, we look at interactions that work with a digital comic on a screen, making static data-comics dynamic by e.g., changing the layout, the visibility of panels, or referencing particular elements. This is designed to complement hand-drawn and computer-generated comic styles equally, by opening spaces for exploration. We hope to increase engagement with the story and data, especially by developing the way that data and narrative presentations relate to each other.

8.3 Authoring Interaction in comics and data visualization

Several digital comic creation systems have been developed for authoring and presenting *interactive* comics, including branching comics with multiple storylines [Tobita, 2015] or non-

branching comics which trace the consequences of audience actions [Soares de Lima et al., 2013]. Making comics tangible, Andrews et al. [2012] built a physical and digital environment allowing an audience to interact with predefined objects in comics drawn on paper above a multi-touch table. Other ad-hoc physical solutions involve creative cutting and folding of paper to support branching stories [E28, E29] (see Section 8.4). However, these tools lack support for the kind of interactions that data practitioners require, such as filtering, leaving a considerable workload for the creator.

For visualization and visual storytelling, researchers have focused on interfaces for non-programmers, supporting direct manipulation and clever data binding methods. Tools have been created for creating visualization from templates [Mauri et al., 2017] as well as creating expressive visualization designs [Liu et al., 2018, Ren et al., 2019, Kim et al., 2017b]. Tools for data-driven storytelling can create infographics [Zhang et al., 2020], static graph comics [Kim et al., 2019], personalized glyhs [Xia et al., 2018] or animated video clips [Amini et al., 2017]. Only few tools provide interactivity through, e.g., navigating a slideshow [Satyanarayan and Heer, 2014], interactive data-driven articles [Conlen et al., 2021, Kwon et al., 2014], showing annotations [Ren et al., 2017], or scrollytelling [flo, Sultanum et al., 2021]. Libraries for creating interactive explorables and articles, such as Tangle [Victor, 2013] and Idyll [Conlen and Heer, 2018] provide for scripting interactions such as sliders and update visual representations based on these values. Tableau also allows the export of interactive visualizations into a specific layout, resembling comics or dashboards. Visual elements in visualizations are highlighted on mouseover and show tool-tips.

To support more flexible workflows, Bigelow et al. [2017] provide a bridge between Illustrator and D3. Our approach presented in this chapter is similar to Bigelow et al. [2017], and Conlen and Heer [2018] in spirit, as we also separate the visual design of comics from the specification of interactivity.

8.4 Interaction for Data Comics

To streamline our exploration and creation of interactive data comics, we define six goals for interaction in data comics. These goals either address shortcomings in static data comics, or open up new opportunities for storytelling, exploration, personalization, and engagement with this medium. We then report on a systematic analysis of interactions in traditional (non-data) comics to collect specific low-level interactions (Figure 8.1) that inform interactions for data comics.

8.4.1 Goals of Interaction

Our goals are informed by extensive discussion among the authors (the authors of the original paper this chapter derived from), our own exploration of creating interactive data comics, our analysis of traditional static and interactive comics and existing static data comics, as well as our own review of the body of work on interaction in storytelling (section 8.2). In defining these goals, we aim to exclude interactions that merely support general engagement without a specific link to data-driven storytelling. For example, general interactions may trigger entertaining animations by moving or shaking elements or panels without any semantics associated with it; objects can be made interactive for the reader to play with it, like a music instrument, or a toy present in a panel scene. As the design space of such playful interactions is virtually limitless, and few studies indicate most effective ones to retain and focus the reader, we opted to scope them out in the present work.

[G1] NAVIGATE—Navigation moves the readers between different parts of the story and provides agency to decide what parts to focus on. Static data comics already enable different levels of navigation through skimming, following the proposed sequence, pausing, focusing only on the text or graphics, or skipping entire panels and pages. Interactions with comics can support or reinforce these different reading experiences, or provide entirely for new modes of navigation through, e.g., hyperlinks, drilling-down into specific panels, flashbacks, or selecting panels of interest. *Interactive navigation in data comics can provide more flexibility to experience a story, follow one's interests, as well as revisit the information in an already read story.*

[G2] DETAILS ON DEMAND—Well known in information visualization, details on demand provide additional information to the interested reader. Static comics and stories are inherently author-driven with the author deciding on the type of information and level of detail in each panel. Moreover, the space for each panel is naturally limited to show multiple panels side by side, which limits the amount and complexity of content that each panel can afford. Interaction can help provide details about, e.g., visual encodings and visualization techniques [Wang et al., 2020], data provenance and transformation [Wang et al., 2021], context, side-stories or other background information. *Designing with* DETAILS ON DEMAND *in mind, allows an author to streamline the storyline while remaining inclusive to audiences with different interests, backgrounds, and data and visualization literacy skills.*

[G3] CHANGE PERSPECTIVE—Author-driven stories usually present a single perspective of the author. This often requires balancing the amount of information with how these information are presented, the author's interests with those of the audience, objective aspects with perspectives of subjectivity, as well as between different stories centered around a particular

data element. For example, a story focusing on geographic data might favor a 'global' holistic perspective over a 'local' more personal one. Interaction can enable readers to select and change main characters, seeing different facets of the data. Switching the main character may filter out panels irrelevant to the new character, add panels directly relevant, and/or update existing visuals to put the spotlight on this new character. CHANGE PERSPECTIVE *allows authors to integrate different perspectives in their story and use complementary, contrasting, or contradictory perspectives as narrative devices [Bach et al., 2018b].*

[G4] BRANCH—Branching introduces forking paths into the storyline, where each path results in a slightly, or sometimes drastically different story. This type of non-linear storytelling provides for a very personal reading experience defined by a set of choices that the reader makes. Branching in static comics leads to structures akin to flow charts. Such comics may prove quite challenging to layout, use a lot of space, and make any single path in the story tree difficult and tedious to reconstruct when looking at the overview (see [E30]). Interaction can support branching while offering a linear reading experience for each path in the tree at any given moment, or displaying two or more paths at once for comparison. *In data comics, branching can be used, e.g., to show different versions and perspectives, contrast data and alternatives, as well as explain possible future scenarios (i.e. "what if?"), or as reactions to the reader's answer to a knowledge question.*

[G5] PAUSE & REVEAL—Gradually revealing information can be used as a device to build tension and force the reader to pause to think, or make a choice. Many web comics have played with gradual reveal through continuous scrolling, or upon clicking on a visual element or panel. Interaction can build artificial barriers for the reader to advance in the story. These range from a simple proof of attention (e.g. mouse over a panel), to an interaction demonstrating the knowledge of the reader or understanding of the prior panels (e.g. multiple choice question), to prompting the reader to make guess or a decision [Kim et al., 2017c].

[G6] INPUT DATA—Data comics and data stories usually contain data provided by the author. However, interaction allows the reader to input or alter data used in the story, decide on manipulations to apply to the existing data, or on parameters for data simulation. For example, providing input mechanisms for readers to enter their calorie consumption or favorite movie can update the panels content for a more personalized, and potentially compelling message. Inputting (personal) data could make data stories more engaging and informative through personalization and explore the effect of different parameters onto the data through dynamic queries.

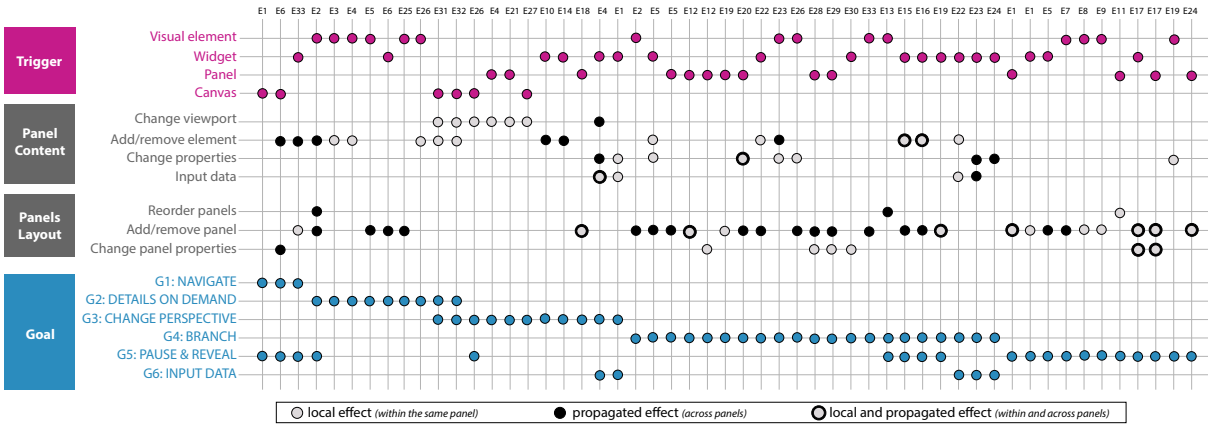


Figure 8.1: Results of the manual coding of interactive comics from our corpus (see list at the end of this chapter). Interactions are defined by a *trigger*, results in visual effects that can affect *panel content*, and/or *panel layout*, to achieve a particular communication *goal*.

8.4.2 Analysis of Interactive (non-data) Comics

We collected 33 traditional interactive comics (found online) to analyze which interactions existing interactive comics already provide, and to inform the authoring process of interactive data comics. Two of the authors (the authors of the original paper this chapter derived from) examined the collection and converged on a code book. One of the authors then proceeded to the coding of the corpus, categorizing interactions into higher level operations (e.g. *add/remove element*, *reorder panel*, Figure 8.1) and relating them to our goals G1-G6.

8.4.2.1 Interaction Analysis—Trigger and Effects

We describe an interaction as having a *trigger* mechanism, that is, how the interaction is activated, and the resulting *effects* of the interaction, that is, what changes occur in the comics. We also found interactions to either impact the *content* within panels, or the *structure* of the comic and its panels. References to example comics are indicated as E1, E2. The full list of example comics is found in the Appendix.

Interaction trigger On the top-left of the Figure 8.1, we list the elements in the comic that acts as the trigger for an interaction. Interactive triggers include (i) a particular *visual element* in a panel, such as an object, or a data point that can be hovered over, clicked, or dragged; (ii) a *widget*, such as a button, a slider, or a text field for the reader to input data or specify a choice, which may be integrated within a panel, or placed on the side; (iii) an entire *panel*, that can be clicked to indicate a choice, or reveal what comes next; or (iv) the whole *canvas*, that the reader can navigate through scrolling [E1], panning or zooming, or even explore in augmented reality

by moving a device around [E19, E32].

Panel Content lists all of the changes that occur within the panels of the comics. These include (i) *change viewport* e.g. when the viewer focuses the story on a particular geographical area, a panel showing geographical automatically recenters on that area [E4]; (ii) *add/remove element* in a panel, for instance, when the reader makes a choice in a given panel, this choice is reflected in the story (e.g. selecting an apple or an orange in [E20] results in the character holding the corresponding fruit in their hand); (iii) *change properties* of existing elements, as for instance updating their color, size or location as the result of a story choice [E4], or putting a halo effect to put emphasis on the visual element; (iv) *input data*, where the content of the panel is updated based on the underlying data that changes, e.g. a line chart [E4].

Panels Layout lists all of the changes that impact the comics structure, i.e., the number and/or layout of panels. These include (i) *reorder panel*, e.g. when panels can be swapped to result in a different story [E12], or moved around the canvas to reveal panels placed underneath [E11]; (ii) *add/remove panels*, e.g., when a choice results in a different ending, panels that are no longer relevant are removed, and new panels are added [E20], or when panels are revealed progressively, either to build up tension and create a surprise effect [E13], or as the result of particular choices [E22]; (iii) *change panel properties*, such as its aspect ratio, or orientation [E12].

8.4.2.2 Findings

The above triggers and effects can be used to enrich the storytelling experience in otherwise static comics. Below, we discuss trends and examples relating to our goals.

Interactions vary in the *scope of their effect*: changes are sometimes contained within the panel where interaction was triggered, or have an effect that propagates throughout the story. Most commonly, we found a choice or object involved in an early panel to be made consistent in the rest of the story [E10,E14]. These are more profound actions that change the course of the story, and therefore the set of panels presented to the reader. We note also instances where data is inputted in a panel, and where this data is reflected throughout the story [E4,E23].

NAVIGATE is typically supported by scrolling, or hyperlinks to jump to a particular place in the story, when it is not an enforced slideshow presentation that resembles flipping pages through. We found interesting instances of aids to navigation in the form of providing a hint of what choices are possible [E20], or what will come next as the result of a choice [E2], when hovering over different interactive panels or visual elements. Other supporting effects include highlighting the panel where the reader is currently at [E13, E30], or even embodying the reader

as a character progressing through panels [E15,E16,E12].

DETAILS ON DEMAND is supported mostly by two types of interaction: showing pop-ups or adding and filtering out elements [E4], as commonly found in visualizations; as well as providing side stories, i.e. branching on a set of secondary panels, to access more details, context, or other perspectives on a point [E6] by adding and removing panels.

CHANGE PERSPECTIVE can be supported through instances where the same character changes their personal perspective, e.g., by changing the viewport through drag-and-drop, or moving and tilting a device in the context of augmented reality [E26,E27]. An interesting instance of a perspective change is by moving a phone device to a particular location on a poster, which then triggers the story of a particular set of characters at the location of focus [E32]. This metaphor is also found in the famous project HERE [E18], where multiple spatio-temporal viewports are materialized as panels around a larger big room, each of which providing a "window" in time of a location in the room.

BRANCH is a popular goal that interaction enables in digital comics. We found many instances where the reader is prompted to make a choice that has repercussion on the rest of the story, by e.g. clicking on a visual element or widget [E22], panel [E20], or indicating the direction of progression [E13]. Other interesting branching instances are the result of personalized input, e.g. the reader writes what a character says [E22], or indicates physical characteristics of a character, which is then used to calculate the outcome of the story [E23].

PAUSE & REVEAL in our corpus uses add/remove elements and panels, for purposes including hiding the result of choices away from the reader to not influence their decision [E13, E28], creating an effect of surprise [E2, E7], prompting the reader to guess a result, or answer a question [E1], or simply because of screen real estate [E5, E17, E26]. Other examples of reveal are achieved through flipping panels [E9], or through changing the perspective with parallax, revealing potentially important/relevant details only visible from a different angle [E26].

We found only a few instances of INPUT DATA, through a range of traditional widgets, that can be hand-drawn and fully integrated within the comic (like the form and slider in [E22], or buttons in [E4]), or computer-generated, like that found in traditional explorable explanations [E1]. An interesting instance of data input we found is randomization [E10]: random generation of content of balloons (in emoticons), resulting in a new story left to the interpretation to the reader, every single time, and even order and content of panels [E14].

8.4.2.3 Summary

There is a rich set of interactions found in traditional comics and many of these operations can directly be used to support storytelling with data comics; hovering elements to emphasize perspectives, adding panels to show details on demand, or posting the reader a question to decide on which panels to show next. The design space in Fig.8.1 serves as a blueprint to apply these interactions to data comics.

8.5 Examples of Interactive Data Comics

Over the course of several months, we engaged in creating several interactive data comics, available on our website: `https://interactivedatacomics.github.io/examples.html`. This section describes two interactive comics to show how interaction has been used for interactive data-driven storytelling. We then report and reflect on our creation process and the lessons learned that motivate our scripting approach discussed in section 8.6.

8.5.1 Example 1: Interactions for Detail-on-Demand, Navigation, and Changing Perspectives

This story presents a dynamic network of **historic alliances** between countries in Europe (Figure 8.2), based on a static data comic [Bach et al., 2016a]. With the goal of informing and educating, this interactive comic offers the option to read the story at different predefined information granularities.

The comic is initialized with two panels (a), a timeline highlighting major events (b), and three version buttons able to change the detail of the comic (c). Clicking on a version button shows the corresponding version of the story: (1) short, (2) medium, and (3) long version (NAVIGATE), adding more events and descriptions at a finer level of granularity. Some panels show small blue *detail* buttons (d). Upon click, these buttons drill down into the respective panel, again providing more information about the respective event shown in that panel (DETAILS ON DEMAND). Hovering over a country label or one of the events in the timeline highlights all occurrences of this element throughout the comic, supporting navigation by indicating what panels are relevant to a particular event or country (NAVIGATE), as well as allowing to follow a single element through the story (CHANGE PERSPECTIVE).

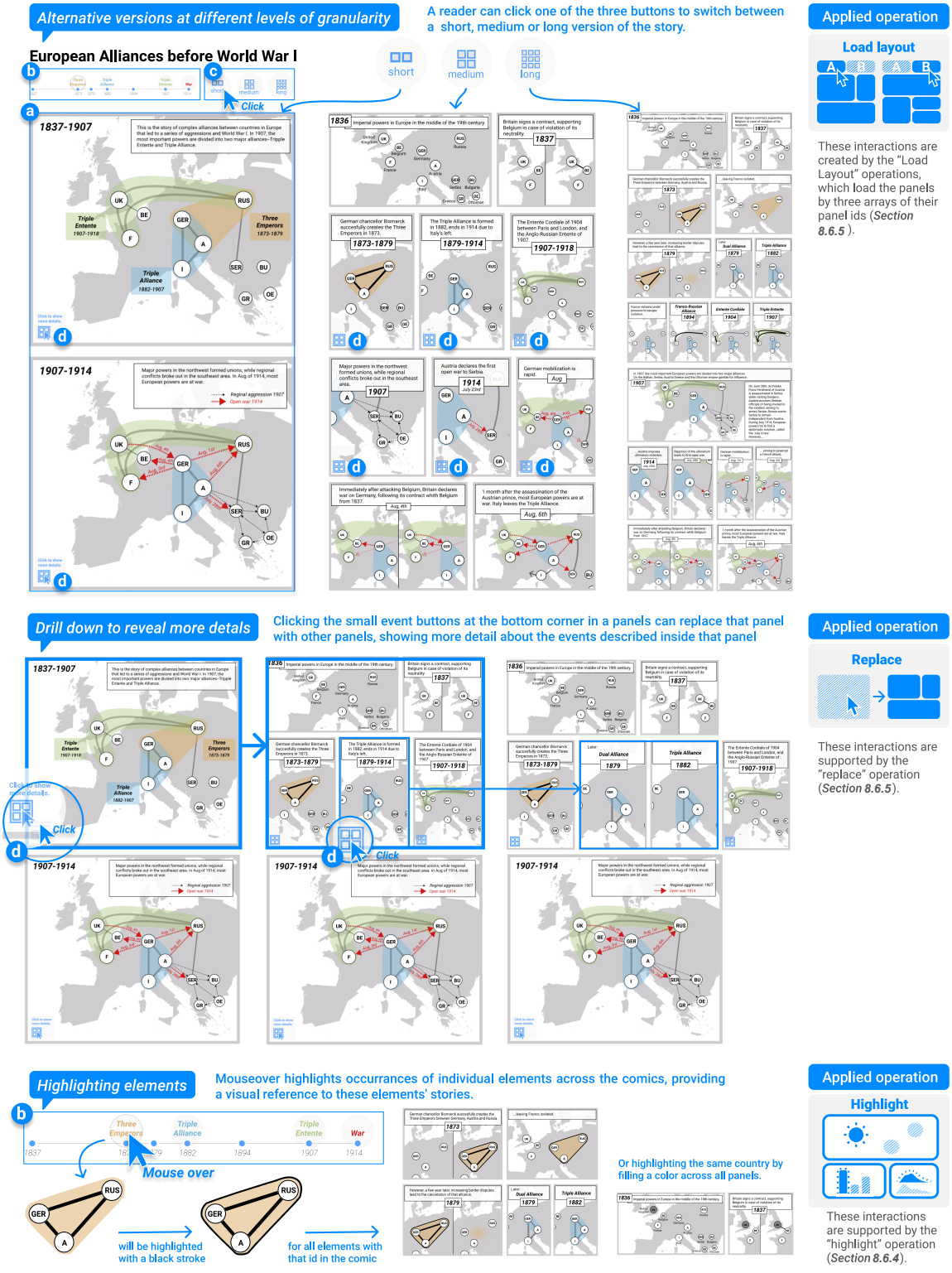


Figure 8.2: The interactive comic of *European Alliances before World War I* presents a dynamic network with a set of countries in Europe, from forming alliances to regressions. With a goal of informing and educating, this interactive comic presents the story of different information granularity (NAVIGATE), provides drill gown to acquire information on demand (DETAILS ON DEMAND), and highlights elements as visual navigation (NAVIGATE).

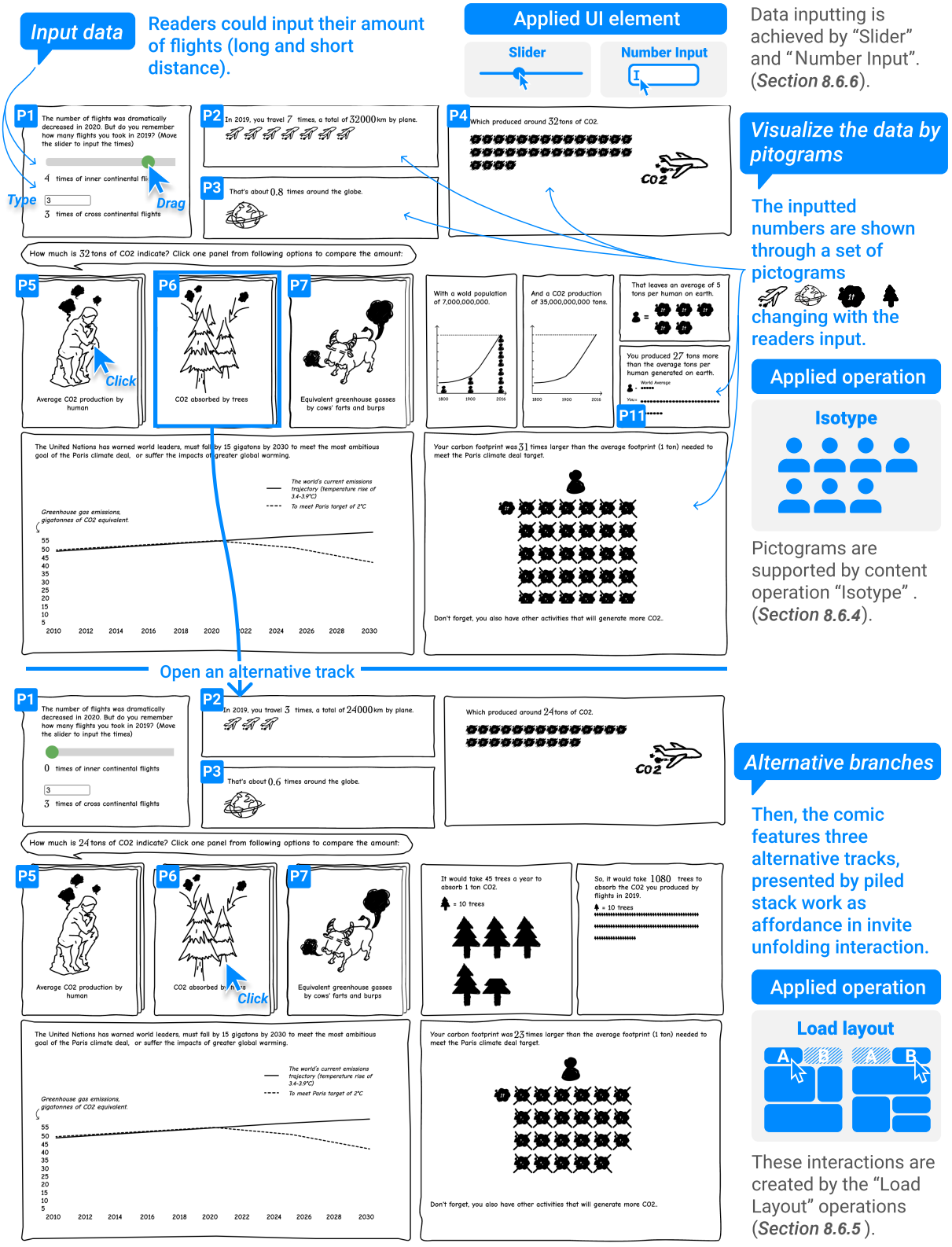


Figure 8.3: The comic of *CO2 emissions* presents viewers with personalized contents by allowing data input, and provides different equivalent comparisons (BRANCH).

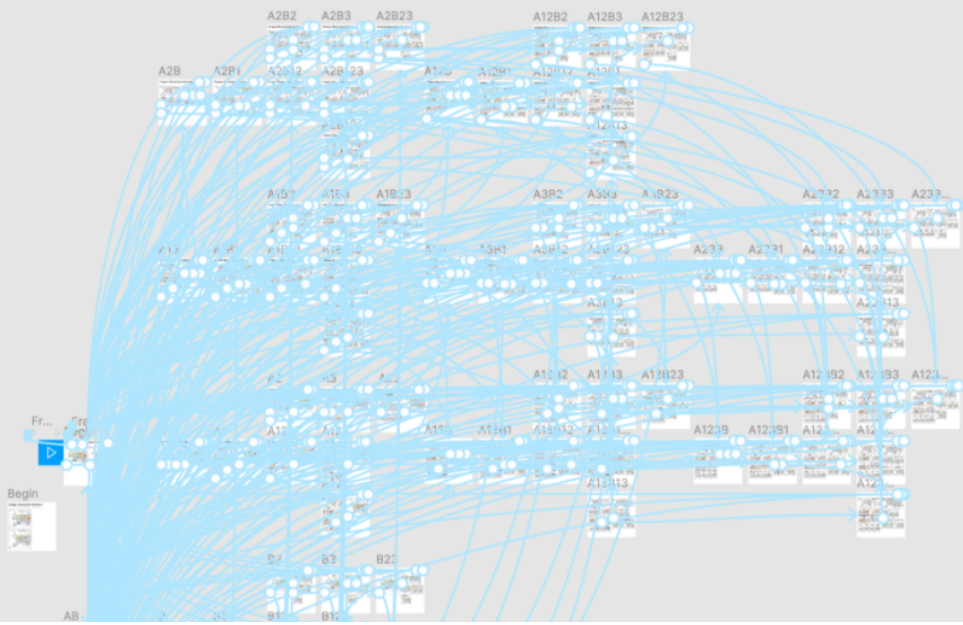


Figure 8.4: Workspace in Figma for the European Alliances comic (Figure 8.2); each small white square is a a full replication of the comic (state) while blue links represent interactive state transitions.

8.5.2 Example 2: Interaction for Data Input and Branching

The example in Figure 8.3 displays the amount of CO2 emissions resulting from personal air-traffic and puts these numbers into context using concrete scales [Chevalier et al., 2013]. The goal is to engage an audience with their personal data (INPUT DATA) and to promote sensitivity for an important issue. The comic allows a reader to input their amount of flights using a widget (P1). Internally, this number is converted into (roughly) the equivalent CO2 emissions, and travel distance. Both numbers are shown through a set of pictograms (clouds and globes, respectively) and change with the readers' input (P2-P4). Then, the comic features three alternative tracks (BRANCH), relating the amount of CO2 emitted to the average CO2 produced by a human (P5), CO2 absorbed by trees (P6) and equivalent green house gasses produced by cows (P7). Panels P5-P7, showing these options, are presented as piled stack inspired by [E11]. Clicking one of these three units adds a set of panels showing the values of the respective comparison. The last row of this comic explains the goal of the Paris climate deal in a static panel and finishes by illustrating the exceed of CO2 of the reader, based on their input.

8.5.3 Authoring Challenges in Using Design Tools

Authoring interactive data comics like those described above requires a wide range of skills and techniques to function, which makes many existing tools not quite the right fit. As traditional comics, they require strong visual design, with fonts, appealing color schemes, layout, and potentially characters and illustrations. As interactive data storytelling media, they require identifying data elements, placing triggers for interactions, and specifying their effects within and across panels.

In our design process, we first tried common design tools such as Figma, and Adobe XD. These interfaces come with sophisticated vector graphics drawing capabilities, as well as features for prototyping interactivity, and live previews. Visualizations and content can be created both inside the tool as vector graphics, or imported as hand-drawn pictures in PNG. Then, a designer can set regions in elements that trigger specific operations which are then prototyped inside the tool, by showing in hiding different frames in elements. While being able to support simple interactions, these no-coding prototyping tools suffer from serious limitations; an abundance of manual specification and an increasing amount of duplicating panels was required for even simple comics and interaction and made it a week-long process to create the first versions of the comics in section 8.7.

The resulting workspace in Figma for the example in subsection 8.5.1 is presented in Figure 8.4. Maintaining complete states of the comics at all points of the interactions required many panel duplication. These number of elements made the workspace challenging to navigate, the general interaction design hard to preview and possibly fixed. In addition, the number of visual assets required affected the performance of the software itself, making it more tedious to author and preview.

Separating the specification of interactions from the visual assets appeared a necessity, to simplify the authoring process but also to enable the creation of rich sets of interactions. In addition, enabling interactions tailored to data comics and triggered when the reader inputs data or select a specific value are missing from generic design tools. For these reasons, we started developing COMICSCRIPT.

8.6 Authoring Interactive Data Comics

To address the issues raised by our initial exploration in section 8.5 and to further explore the possibilities for interactive data comics, we developed COMICSCRIPT, a lightweight, declarative language for specifying comics and interactions over them, with particular support for data

comics. COMICSCRIPT is our approach to balance and combine aspects from graphical user interfaces and the power of coding interaction. COMICSCRIPT is based on JSON and can be edited and interpreted by an online editor, and the resulting interactive comics can be exported and linked into any website. COMICSCRIPT captures most of the interactions set out in our design space (section 8.4) through a set of *operations* that change content, layout, structure and appearance:

- **structure operations** manipulate structure and layout of a comic, adding or replacing panels or **specifying panel layout** to make wholesale changes to the layout of all or part of the comic. This includes navigation such as jumping the reader to a specific panel.
- **content operations** use **classes of objects** to **highlight** and group visual elements across the comic, or apply Isotype techniques to indicate the value of variables.
- Techniques for **data input** allow users to input numbers through text or sliders and compute **variables and formulas** that can be used in the content or to create **conditions** on whether other operations can be triggered.

In the remainder of this section, we first explain the steps involved in creating interactive comics and then explain how COMICSCRIPT supports scripting interactive comics.

8.6.1 Creating Comics with COMICSCRIPT

Our process of creating comics has two stages. In the **preparation phase**, a designer specifies each individual panel and its content. This phase includes thinking through the comic layout, as the designer needs to decide on a specific layout of panels, i.e., the size and their order to achieve a meaningful reading order. Here, they can use their usual workflows: a designer could create visualization contents and panels inside vector graphics tools, import hand-drawn images, or rely on existing visualizations in both pixel and vector format. The only restriction is that a single SVG or PNG file should be produced for each panel in the comic, and they should be provided at a public URL so that the web editor can load them.

A COMICSCRIPT comic starts with definitions for each panel, containing a unique ID and a URL for the image (subsection 8.6.2). An initial layout is also specified, using the panel IDs in a nested structure to describe rows and columns (subsection 8.6.2). At this point, the comic can be displayed as a static webcomic, while interaction is added next.

In the **scripting phase**, the designer defines the list of operations that can be applied to the comic. Operations are attached to specific SVG elements through their unique IDs and parameterized in the script. Operations such as adding panels or setting a new layout refer to

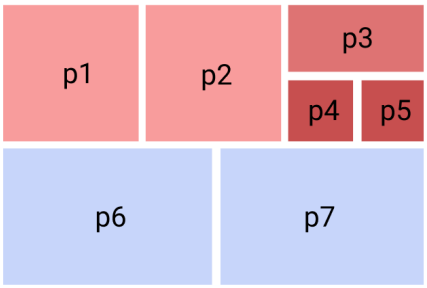
the panels through their IDs. The set of **structure** operations (subsection 8.6.5) works equally well with PNG or SVG files, and allows a large degree of flexibility with panel manipulation. However, if the graphics have been created as vector-based SVG files, a wider range of interaction is possible, as the individual elements in the graphic can be addressed. Vector elements can be used as triggers for operations, e.g., functioning as buttons, or as the targets for effects, such as being added or removed, or highlighted. Objects with the same ID will be applied the same effects, allowing for cross panel interaction, such as brushing to highlight related data. Vector elements with IDs also allow the integration of text input, sliders, isotype pictograms and other types of interaction (subsection 8.6.4). The remainder of this section explains the individual components of ComicScript, and how they come together to create interactive comics. Our set of operations represents a careful balance of low-level operations as they are described in the design space in subsection 8.4.2 (e.g., append a panel) and high-level operations that subsume a set of low-level operations (e.g., load layout). More information and a detailed tutorial are included on our website.

8.6.2 Panels and Layout

ComicScript loads panels from a URL and assigns them a string ID, which is used throughout the specification to refer to this panel.

```
{"panels": [{ "id": "myPanel", "url":"panel1.svg"}]}
```

Since the existence and order of panels can vary over the course of the interaction, ComicScript provides a flexible way to define layouts. With the width of each panel given by the SVG or PNG image, a layout in ComicScript can be specified as a nested structure (e.g. panels organized in rows, organized in pages). Alternative layouts can be specified and loaded upon interaction, resulting in rearranging, adding or removing panels. Layout are thus expressed




```
{"layouts": [{
    "name": "init",
    "panels":
    [["p1","p2",["p3",["p4","p5"]]],
    ["p6","p7"]]
}]}
```

as nested arrays, the first corresponding to a row in a comic, the second to a column in this

row and so on (see example above). Panels within a row or column are laid out side-by-side, separated by a gutter which size is specified by the `gutter` parameter.

8.6.3 Data and Variables

COMICSCRIPT allows operations to refer to constructs and data: classes and variables. **Classes** group visual elements in the comic into groups, which can be used to highlight elements and refer to groups of elements. Classes are only possible with SVG panels since they require the referred elements have unique IDs in the file. IDs can be the same across panels, e.g., the first bar can always be called `bar1` across several panels with a bar chart.

```
{"classes":{
   "name": "countries",
   "elements":["france", "germany", "uk"]
}}
```

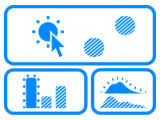
Variables store numerical values which in turn can be shown inside a text element through a place holder that has the same ID as the variable, or used to render a number of ISOTYPE like symbols. Variables can be obtained from data, or through user input, e.g., using a slider.

```
{"variables":[
   {
     "name": "movement","value": 0
   },{
     "name": "totalkm",
     "value": 0,
     "what": [ "movement","*1000"]
   }]
}
```

The optional `value` field specifies the default value if the variable is not set. Calculated variables are created by functions, expressed through a `what` attribute. Variables can come directly from data or input via input elements (subsection 8.6.6).

8.6.4 Content Operations

According to our design space in subsection 8.4.2, content operations define which elements are shown inside panels and how they are shown. COMICSCRIPT currently provides three content operations.



Highlight changes the appearance of all visual elements with the same ID across all panels (e.g., panels or SVG elements) on mouseover. Highlighting can be used as a visual reference when the author wants the audience to look back/forward at a certain panel or show that an element is the same across the panel (CHANGE PERSPECTIVE, NAVIGATE). An example of highlight is shown in Figure 8.2-bottom). The appearance of a highlighted element is specified through its CSS attributes `scale`, `stroke`, `color`.

Layers enable showing and hiding SVG layers and their elements. This operations could be used for presenting DETAILS ON DEMAND and linking texts and its visualisations NAVIGATE. Figure 8.5 shows an example how layers can be used to show temporal data as the users clicks dates.

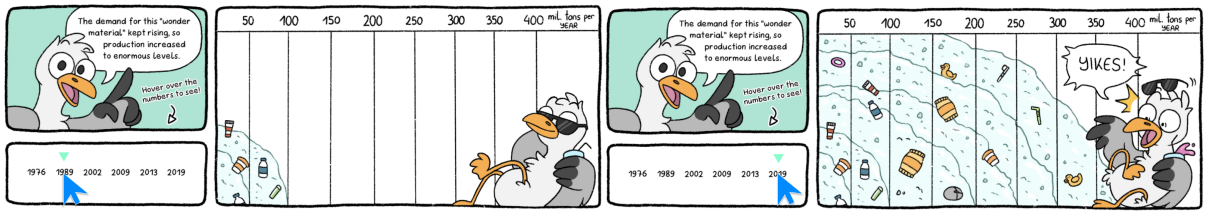
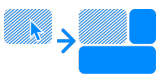


Figure 8.5: Using layers, hovering over different time on the timeline triggers different layers.

Isotype aims to reduce the difficulty of understanding abstract content and to support INPUT DATA through the user. The operation creates one or more ISOTYPE like pictograms inside a specified element (i.e., `TreesPlaceHolder` in Figure 8.3) and binds the number of icons to the value of the `variable`. The size of each icon could be adjusted in `widthIcon`. The icon is loaded from the SVG specified by `url`.

8.6.5 Structure Operations

Structure operation change the set of shown panels and potentially their order. COMICSCRIPT provides the following structure operations: *append panel, load layout, replace panel*, and *pan and zoom*.



Append adds one or more panels after a specified panel. This operation could be used for different narrative purposes, e.g., BRANCH; providing DETAILS ON DEMAND by drilling down from this panel (Figure 8.2-center); revealing the answer for a question raised in the panel; or PAUSE & REVEAL. This operation does not remove or replace any panels. If the creator wants an operation to remove any panel and load new panels, 'replace' or 'Load layout' should be applied instead. If multiple panels are inserted, the same notion as in layouts is used: e.g., `["p3",["p4","p5"]]`.

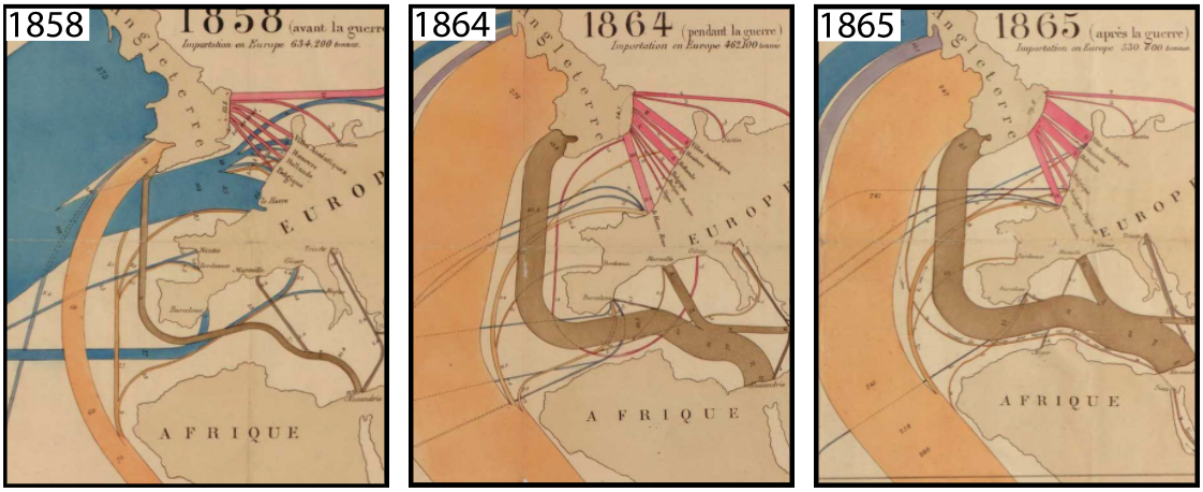
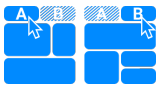
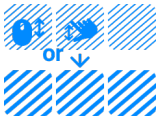


Figure 8.6: Example for a pan and zoom operation: panning and zooming in one of the panels with a map synchronize the other panels so that the same detail of the map is shown in all three years.

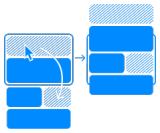
Replace removes a panel to replace it with one or more (`newpanels`). This operation can be used for replacing panels presenting overview with details or replacing the following narratives by selecting different data set as inputs (CHANGE PERSPECTIVE). Replacing an overview with details (DETAILS ON DEMAND) (e.g., clicking the button (d) in Figure 8.2).



Load layout enables to display a fresh set of panels in a specific layout, removing any panel specified by the `after` attribute. This high-level operation can be used similar to a menu on a website (NAVIGATE), which shows alternative versions, e.g., in length, visual style or narration, (Figure 8.2). Similar to `append`, `loadlayout` could also be used for presenting DETAILS ON DEMAND, CHANGE PERSPECTIVE and BRANCH.



Pan and Zoom creates a simple pan and zoom behavior for a single panel, and can synchronize pan and zoom across multiple panels (e.g., in Figure 8.6). Pan and zoom help to achieve the goals CHANGE PERSPECTIVE and DETAILS ON DEMAND.



Jump brings the user to another panel, similar to anchors in HTML. This operation is designed for non-linear reading by NAVIGATE, and to provide visual navigation in a web comic when a pointed panel is not visible due to limited screen real estate.

`[x > 10]` **Conditions** can be attached to operations to express constraints when an operation is triggered. This operation allows for personalization, e.g, through INPUT DATA. For example, depending on the user's input in P11, Figure 8.3 the comic tells *tons more than*, *tons less than* or *equal to* when comparing that user's data to other values.

8.6.6 UI Elements for Data Input

UI elements can be added into SVG panels, replacing an element with a specific ID. Values from these UI elements are then mapped to variables and used to, trigger conditions or show ISOTYPE symbols. COMICSCRIPT currently supports a slider and number input, both demonstrated in Figure 8.3.

8.7 Creating Interactive Data Comics with Designers

We asked designers, illustrators and visualization practitioners to create interactive data comics with our tool. We wanted to gain insights on many aspects of the creation process as well as improve COMICSCRIPT. We gathered insights on the workflow people follow to craft interactive data comics, their biggest challenges, the role of COMICSCRIPT in the creating process, the purpose of interaction and the corresponding selected interactive operations as well as identifying missing interactions and operations to iterate on COMICSCRIPT. Respective tutorials and examples can be found on our website.

8.7.1 Procedure

We conducted a creation workshop consisting of a series of live sessions to help participants create their own interactive comics using COMICSCRIPT. First, we ran a tutorial session, introducing the concept of data comics and interactive data comics. We showed examples of data comics, interactive data and non-data comics. Then we used a simple example to explain the workflow we followed (section 8.6) including how to create and export individual comic panels with a design tool (Figma) and how to add interactions to the comics using our scripting approach. Participants attended the tutorial live or could consult the recorded version. We offered 2 hours of daily face-to-face online drop-in sessions for answering questions about designing and using COMICSCRIPT. Participants volunteered for an individual 15-minute semi-structured interview after the completion of their comics. During the interviews, we asked five questions followed by an open discussion. Questions asked for *1)* the desired intentions of interaction for data comics, *2)* inspiration and ideation of interactions, *3)* creation process and challenges with COMICSCRIPT, *4)* possible future interactions, and *5)* the potential of interactive storytelling.

8.7.2 Participants

Six participants (four from Europe, one from Canada and one from Asia) completed their interactive comics and attended the interview. The participants were: a professional illustrator for 15 years (*Illu*), a postgraduate student in digital media with 6-year experience of creating illustration (*IlluDigt*), a computer science postgraduate student with 15-year experience working as an illustrator (*IlluDigitComp*), a postgraduate student in design and data science (*DesignData*), a doctoral student in creative art and visualization (*ArtViz*), and a team of data storytellers from the company Gramener [Gramener, 2019] (*DataStr*) with experience in creating data comics and interactive data comics by programming.

8.7.3 Qualitative Findings

We gathered feedback during the creation process, and the interview. We organize insights around three themes: creation process, COMICSCRIPT, and participants' perspective on the value of interaction.

Creation process. Participants worked at their own pace and contacted the instructor occasionally, mostly towards the end of their creation process. Once selecting a topic and gathering data for their story, participants devoted the largest portion of their time to crafting the visuals (e.g. elements and panels) of their comics (seven to ten days). For this stage, they used graphics editing tools such as Adobe Illustrator, Photoshop, Figma, INKSCAPE (`https://inkscape.org/`) and Microsoft PowerPoint. Participants reported thinking of the interactions as they designed the comics and crafted the visuals. They integrated the interactions using COMICSCRIPT over one to three days.

The overall comments suggest that all participants appreciated the final product. This was particularly salient for creators with no coding experience: *"I have all this inspiration imagination and creative ideas but I've never been able to transfer them into code because it's a completely different realm, so what this does is it allows you (a creator) to take your creative imaginative processes ideas and concepts put them into an already constructed editor and then feel proud of yourself."* (*ArtViz*). Several participants gravitated towards *"different perspective"* (*Illu*) (CHANGE PERSPECTIVE) and non-linear comics (BRANCH) *"like Chris Ware [in which] you can almost just choose a panel on a page and then see where it leads"* (*IlluComp*).

Participants reported that spending the **majority of their time in graphics editing software** to create the story and design the comic panels. To create different layers or branches, they needed to make more images and variations than it presented. Also, participants **switched between tools**, e.g, taking raster images to vector editors to add IDs if there were buttons or

placeholders in the panels, and making sure the size of the image was not changed across tools (*IlluDigt, DataStr*). After the images were exported, participants started editing from one of the examples on our website, since *"the editor already had like a pretext in it and because I could see what your functionality did and with the what your process was I could just change parts of that code and I didn't have to know too much as a user."* (*ArtViz*).

COMICSCRIPT usage. Participants completed their comics **with little to no help** from the facilitator. Overall, they used nine out of eleven operations in the comics they created. Most participants used operations dealing with the panel layout (*Load layout*, *Replace* and *Append*), as well as content operations across panel (*Highlight*, *Layers*). Half of the participants also used operations enabling readers to modify the data (*Slider*, *Condition Number input*, and *Isotype*). COMICSCRIPT definitely involved **a learning curve**: *"Once you understand it then it's totally fine"* (*IlluDigt*). After a few explorations using the code samples provided, participants were able to implement their ideas: *"I always felt like I knew what I wanted to do and I knew how to do it"* (*IlluComp*). Two participants expressed their enthusiasm for a user interface layer to be added on the current version, enabling them to drag-and-drop the interaction of their choice directly on top of the comics elements. The JSON language was commented as *"a very refined code for a specific task"* (*ArtViz*), and *"really structured"* (*IlluDigt*). The most challenging aspect reported by participants was the **low-tolerance to common syntax error** of COMICSCRIPT.Four participants suggested **extending current operations**, for example, to enable panel replacement when using a slider. Two participants wanted **a finer control** over operations such as resizing or translating elements, which is implemented as low-level operations in COMICSCRIPT but was not surfaced to them at that point.

Goal of interaction. With regards to the **potential of interaction** in data comics, participants indicated that providing navigation (NAVIGATE) and branching (BRANCH) enabled conveying information from different perspectives (CHANGE PERSPECTIVE), e.g., *"[Readers] can just dive right into it click on something see the connection and it's way easier to get"* (*IlluDigt*). Participants commented that interactions made data more engaging, *"get the curiosity of people"* (DataStr), *"with interactivity you can kind of be[ing] inside of it"* (*IlluDigt*), and potentially being used *"as weekly reports replacing dashboard"* (*DataStr*) or in news article to enable audience to *"think beyond just the passive presentation of imagery"* (*IlluComp*). A participant suggested that crafting interactive comics with COMICSCRIPT could both **foster creativity and [enable] learning programming**, since this activity *"introduces people to code but also within the creative process, creativity and data can be pushed apart so it's kind of merging them together in this very nice safe space"* (*ArtViz*).

8.8 Discussion and Future Work

Crafting interactive data comics involves challenges at multiple levels of the creation process. For several of our workshop participants, the experience of using COMICSCRIPT was their first attempt to create dynamic, computationally mediated material. Challenges occur at multiple levels: from the design and conception of the interactions and non-linear narratives, to the creation and organization of the collection of assets composing the comics, to the implementation of the interactive medium. While the effort is substantial, our workshop revealed that creators see tremendous opportunities in incorporating interactions in data comics, making it worthwhile.

Participants' feedback also suggests that integrating interaction adds many design considerations to the design of comics—what comics structures can benefit most from interaction, how to make interactivity discoverable—and presents a set of new challenges—designing multiple combinations of content and panel layout and keeping track of the state of the comic at a given point. To provide guidance and potential solutions for these design considerations, further research is needed to evaluate use and engagement with interaction in data comics.

COMICSCRIPT: a lightweight language to create interactive data comics. COMICSCRIPT provides a scripting mechanism to add interactivity to static comics. It serves as a conceptual middle ground between traditional interactive prototyping tools, and full-fledged programming languages. Compared to interactive prototyping tools, COMICSCRIPT facilitates experimenting with interactions and scaling in number and complexity of the comics (e.g., exploring non-linear data comics). Compared to classic web programming combining HTML, CSS, and Javascript, COMICSCRIPT lowers the complexity of the code and thus lowers the learning curve and saves time to professional developers. COMICSCRIPT provides a core set of interactions that designers, illustrators and creators with minimal or no coding skills can copy and paste and apply to sets of elements and panels crafted in graphics design tools. Our goal with COMICSCRIPT is to strike a balance between approachability of the language and interaction capabilities added to static comics.

Extending COMICSCRIPT while preserving its simplicity. Participants suggested many extensions to COMICSCRIPT: from low-level fine control of interactions (e.g. resizing or rotating elements in panels) to higher-level more advanced layouts (e.g. nesting or overlaying panels). Other desirable features include more sophisticated animated transitions and the creation of visualizations, annotations, data comics design patterns [Bach et al., 2018c] as well as enabling incorporating different affordances to convey interactivity. Animated transitions can help readers understand changes to the comic following an interaction. Creating visualizations, on the other

hand, could be done by adding specific directives to the language, e.g., by calling on D3 or Vega-Lite [Satyanarayan et al., 2017]. Creating other content, such as text, characters or annotations, could happen through similar approaches, as partially explored elsewhere [Gramener, 2019]. Adding different affordances to suggest interactivity is a research direction in itself [Boy et al., 2016]. The difficulty with these extensions is to balance the power of the language *vs.* its simplicity. While adding lower-level operations (e.g. hide or resize elements) is trivial, adding more complex functionality may necessitate writing more complex specifications even for simpler examples. To avoid re-implementing functionality and avoid frequent switching between design tools, COMICSCRIPT could be integrated with other visualization and comic design tools.

Towards easier authoring of interactive data comics. COMICSCRIPT is a first step towards understanding what interactions matter most to authors and constitutes a probe into the process they are following when creating them. Still, COMICSCRIPT requires designers to learn a new syntax, although it is arguably simpler than learning full-fledged programming languages. To further lower the effort to create interactive data comics, future work should explore graphical user interfaces that can generate COMICSCRIPT code from directly manipulating the graphical elements in comics and their interactive operations. Perhaps one of the most pressing issues in streamlining the authoring process is to facilitate the creation, reuse and organization of the many graphical and interactive assets necessary for interactive data comics. Observations and comments from our study participants confirm that interactive data comics require a much larger set of visual assets than traditional comics. Managing these assets and generating the variations involved in each interactive state is arguably the main bottleneck for the creation of this new medium and which should be aided through future user interfaces.

Eventually, since our approach is agnostic of panels' content, it can be extended to other genres of data-driven storytelling such as infographics, flow diagrams or data-driven articles; sets of panels can simply be defined as individual graphical components of these artifacts.

Beyond tools to craft interactive comics, we need to support a change in practice. Authors and designers need to adapt both their mental model and their creation workflow to design and execute interactive data comics. The learning curve is not only about mastering tools (e.g. COMICSCRIPT) to craft them, it largely lies in adjusting the mental model to design them: understanding what is possible, and then what is desirable. A key enabler is certainly to provide inspiration and examples to draw from—participants relied on examples created with the system and existing interactive comics to develop a sense of what is possible. Additional material to support and inspire the design of interactive comics is certainly critical for the future.

8.9 Conclusion

In this chapter, we bring interaction to static data comics to enable non-linear storytelling, personalization, switching different layouts, and levels of details. We explore interactive data comics by learning from traditional interactive comics and creating examples. We present a set of *operations* such as loading and removing predefined panels and inputting values to support the narrative goals of the comic creator. To support designers to rapidly build web-based interactive data comics, we introduce COMICSCRIPT, a light-weighted specification language based on JSON. We conducted an interactive data comic creation workshop with six professional illustrators, designers, and data comics enthusiasts. In the workshop, we found participants could build their interactive data comics with our approach with little to no help from instructors. This chapter is built on our understanding of static data comics from previous chapters, aims to extend the usage of data comics to more possibilities. We hope our initial exploration towards understanding and creating interactive data comics could inspire practitioners and researchers to create more examples and creation tools.

Catalog of interactive comics

[E1] N. Case. How to Remember Everything Forever-ish. 2018. https://ncase.me/remember/

[E2] R. Chang. What the Blob. Undated. https://rosalind-chang.myportfolio.com/what-the-blob

[E3] M. Cisneros. Is That True?. 2017. https://public.tableau.com/profile/mikevizneros#!/vizhome/IsThatRight/IsThatTrue

[E4] Gramener. Yikes! 6ft please. Undated. https://gramener.com/nyc311/

[E5] Goodboy Digital. Tell me your secrets. https://www.goodboydigital.com/case-study/tell-me-your-secrets

[E6] M. Huynh. The boat. Undated. https://www.sbs.com.au/theboat/

[E7] M. Kühni. CLAMS II. Undated. https://abwaesser.net/comic/clams-ii/#drop

[E8] M. Kühni. Coming of age. Undated. https://abwaesser.net/comic/coming-of-age/

[E9] M. Kühni. Heidi. Undated. https://abwaesser.net/comic/heidi/

[E10] M. Kühni. New speak. Undated. https://abwaesser.net/comic/new-speak/#

[E11] M. Kühni. Visiting the aquarium. Undated. https://abwaesser.net/comic/visiting-the-aquarium/

[E12] Loveshack Entertainment. FRAMED Collection. 2018. http://framed-game.com/

[E13] D. Merlin Goodbrey. A Duck Has An Adventure. 2012. https://www.youtube.com/watch?v=YBRhf2jMHvM

[E14] D. Merlin Goodbrey. Hidden factions. 2000. http://e-merl.com/hyper/hapf/hf1.htm

[E15] D. Merlin Goodbrey. Icarus Need. 2012. https://www.youtube.com/watch?v=tuzaKLowWTA

[E16] D. Merlin Goodbrey. The Empty Kingdom. 2014. https://www.youtube.com/watch?v=9vKuLap55CM

[E17] S. McCloud. The Right Number. 2003. http://www.scottmccloud.com/1-webcomics/trn-intro/index.html

[E18] R. McGuire. HERE. Undated. https://screendiver.com/directory/here-by-richard-mcguire/

[E19] Monde Binaire. Hello World !!!. 2012. https://transmii.com/project/monde-binaire/

[E20] C. Muckenhoupt. Interactive comic prototype. https://muckenhoupt.itch.io/interactive-comics-prototype

[E21] R. Munroe. xkcd: Click and Drag. https://xkcd.com/1110/

[E22] R. Munroe. xkcd: Lorentz https://xkcd.com/1350/

[E23] R. Munroe. xkcd: Throw calculator. https://xkcd.com/2198/

[E24] R. Munroe. xkcd: xkcloud https://xkcd.com/1506/

[E25] R. Munroe. xkcd: Zach Weiner (SMBC) https://xkcd.com/826/

[E26] Netwars. The Butterfly Attack. 2014. http://www.netwars-project.com/novels/graphic-novel

[E27] PLASTIEK. RRR2. 2018. http://plastiek.com/RRR-2/RRR-2.html

[E28] J. Shiga. How To Make Your Own Interactive Comic. 2010. https://www.youtube.com/watch?v=A9Er7kmcPcI

[E29] J. Shiga. Numberphile. Paper Caluculator. 2016. https://www.youtube.com/watch?v=siawhQBRC8I

[E30] J. Shiga. Meanwhile. https://www.zarfhome.com/meanwhile/

[E31] Sutu et al. Modern Polaxix. 2014. https://screendiver.com/directory/modern-polaxis/

[E32] Transmïi Studio. Memory Catcher. 2014. https://screendiver.com/directory/memory-catcher/

[E33] Turbomedia. Coup de Pompe. Undated. https://turbointeractive.fr/coup-de-pompe/

Chapter 9

Conclusion and Future Work

This thesis is primarily motivated by two questions: *Is data comic an effective genre for data-driven storytelling* and *How to create effective data comics*. During our exploration progression, these two questions are interacted and reinforced each other. To better understand data comics, we began with a study to investigate how reading guidance and text-picture-integration affect the understandability, memorability, and engagement in data comics and infographics. The study results inform design guidance to be used in data comic creation workshops. Further, the workshop enhances our understanding of the importance of using and explaining data visualizations in data comics. This challenge motivates us to provide Data Visualization Cheat Sheets for interpreting visualization techniques that are appropriate for applying in comics. Moreover, Data Visualization Cheat Sheets are then also applied in data comics as solutions for reporting quantitative studies in visualization and HCI. With complex data and stories such as scientific reports, being able to actively access data as required is essential for effective reading for an audience, the progress of understanding and creating interactive data-driven stories with data comics are also made in parallel. The Data Comic Creation Model presents the relation among essential elements of data comics and a methodology of creation. Data comics are an effective genre for data-driven storytelling. However, creating compelling data comics requires multiple skills and careful thinking of design trade-offs. This thesis provides workshop procedure, materials, design supports, and guidance to foster the use and the creation of data comics.

In this chapter, I present a summary of this thesis by highlighting the contribution of individual chapters. This chapter focus on how the contributions address the initial seven research questions raised in Chapter 1 and outline planned future work with open questions.

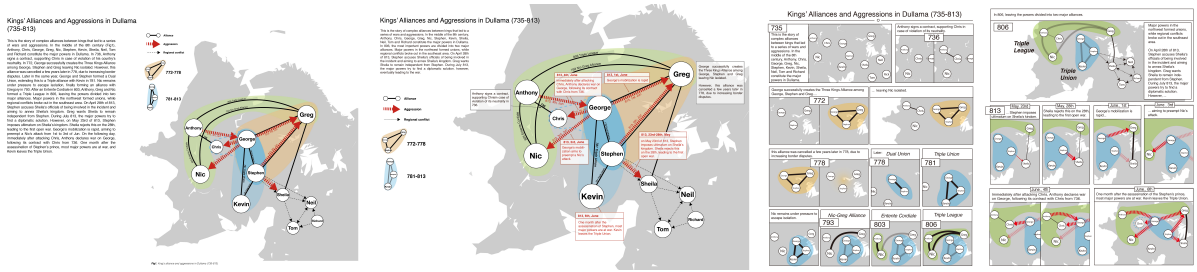


Figure 9.1: Samples of ALLIANCES in the lab study of Chapter 4. ILLUSTRATEDTEXT (left) ; INFOGRAPHIC (center); COMIC (right).

9.1 Summary and Contributions

Chapter 3: Towards a Data Comic Creation Model This chapter presents Data Comic Creation Model to foster understanding and create data comics. Inspired by the five plane model of designing user-centered products [Garrett, 2010], Data Comic Creation Model is a structured approach to helping novices and professionals in design, illustration, data science to create storytelling, especially, to describe the relationship among the essential components and creative methods of data comics. The model has five layers, from the bottom to the top reflects a typical workflow of creating data comics: I) Communication Objective, II) Message, III) Narrative structure and Explanatory Visualizations, IV) Visual Narrative Integration, and V) Graphic Appearance. This model contributes not only a summarised instructional approach for creating data comics, but also evoke future research questions, tools, pedagogy for creating data comic, and potentially other visual data stories. It generalizes well beyond data comics and will help educators, novices, and professionals create stories.

Chapter 4: Comparing Effectiveness and Engagement of Data Comics and Infographics

Inspired by how people read and use comics, data comics are initially static format for presenting data stories, we begin with asking *RQ1) Are data comics more effective compared to other data-driven storytelling formats?* and *RQ2) What are the merits and drawbacks of data comics?* To answer these two questions, we conduct a lab study including a series of tasks that focus on participants' reading styles, comprehension and memorability among ILLUSTRATEDTEXT, INFOGRAPHIC, and COMIC (Figure 9.1), and an in-the-wild study that focus on how pedestrians engagement with data INFOGRAPHIC and COMIC (Figure 9.2).

Based on a variety of collected data including questionnaires, interviews, think-aloud protocols, explanations, memory-recall, and visitor observation, we found that data comics are seen as more fun, help readers to stay focused, and are overall more engaging. Quantitative results

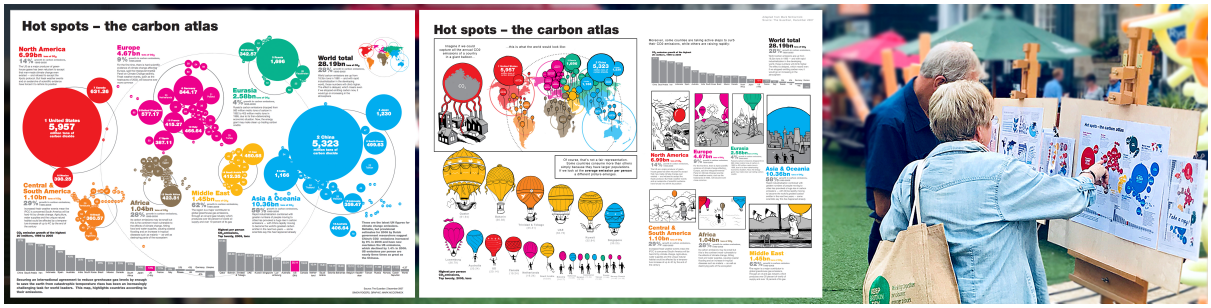


Figure 9.2: Two stories were presented in the wild as an INFOGRAPHIC and a COMIC for an empirical observation study conducted with pedestrians, measuring reading time, interactions (i.e. pointing and talking) and opinions as evidence for engagement and enjoyment (Chapter 4).

suggest that data comics improve understanding for most cases, and were generally rated high for enjoyment and engagement. In addition, subjective feedback collected gives a richer picture of the respective merits and drawbacks of each format. The results inform the design of future data comics and communicative graphics.

Chapter 5: Teaching Data Visualization and Storytelling with Data Comic Workshops The promising results from Chapter 4 motivate us to explore more on how to create effective data comics, in particular, *RQ3) What is the process of creating data comics?* This chapter offers a detailed protocol of how we ran the workshop, so that others may reproduce it. We also present reflections and possible improvements for future scenarios. In summary, we found that the illustration students' confidence in drawing led to engaging and inventive imagery and narratives. Once they had found their data and wrestled with the storytelling they really engaged and explored the key elements of data comics and began to touch on data journalism and info graphics. We found that students had challenges in 1) the data visualization stage, in selecting visual variables or types of visualization techniques, and elaborately explaining them clearly to audiences; 2) the story-boarding stage, especially when students did not have a concrete idea of a story, and a simple method to experiment design possibilities; 3) Working with design patterns to create data comics, participants found it hard to understand design patterns given the short amount of time. Overall, our updated workshop design led to better results compared to the two previous workshops. Our reflections led us to think further on data comics authoring and teaching methodology and provide guidelines for holding data comics workshops and data comics pedagogy in the future.

Chapter 6: Cheat Sheets for Data Visualization Techniques Data visualization techniques are essential to communicate insight from data and deliver evidence in data-driven storytelling. However, we found in Chapter 5 that data visualizations found not easy to create, interpret

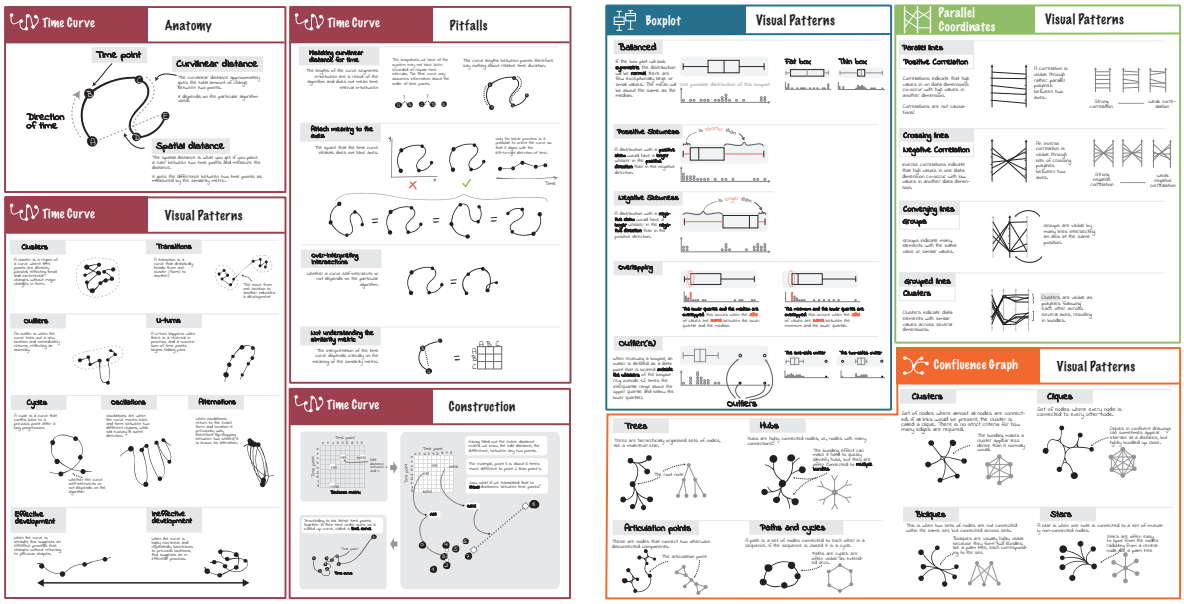


Figure 9.3: Examples of customized cheat sheets combining several aspects of visualizations. Left: example of look-up summary sheets for the time curve visualization showing *anatomy*, *visual patterns*, *construction*, and *pitfalls*. Right: Cheat sheet showing *visual patterns* for several techniques: boxplots, parallel coordinates, and confluent graphs.

properly, especially for data and visualization novices. To answer *RQ4) How to effectively explain data visualization techniques in data comics*, Chapter 6 introduces the concept of "cheat sheets" for data visualization techniques, a set of concise graphical explanations and textual annotations inspired by infographics, data comics, and cheat sheets in other domains. Cheat sheets aim to address the increasing need for accessible material that supports a wide audience in understanding data visualization techniques, their use, their fallacies, and so forth. We have carried out an iterative design process with practitioners, teachers, and students of data science and visualization, resulting in six types of cheat sheets (*anatomy, construction, visual patterns, pitfalls, false-friends, and well-known relatives* (e.g., Figure 9.3) for six types of visualization, and formats for presentation. We assess these with a qualitative user study using 11 participants that demonstrate the readability and usefulness of our cheat sheets.

Chapter 7: Data Comics for Reporting Controlled User Studies in Human-Computer Interaction Communication is crucial in academic research for both experts and non-experts, to precisely deliver information with limited time and attention from the audience is a challenge. For example, scientific studies from the field where the author of this thesis from, human-computer-interaction and data visualization, are typically involved controlled experiments. Reports details hypotheses, conditions, data collection, experimental setup, procedure, analysis,

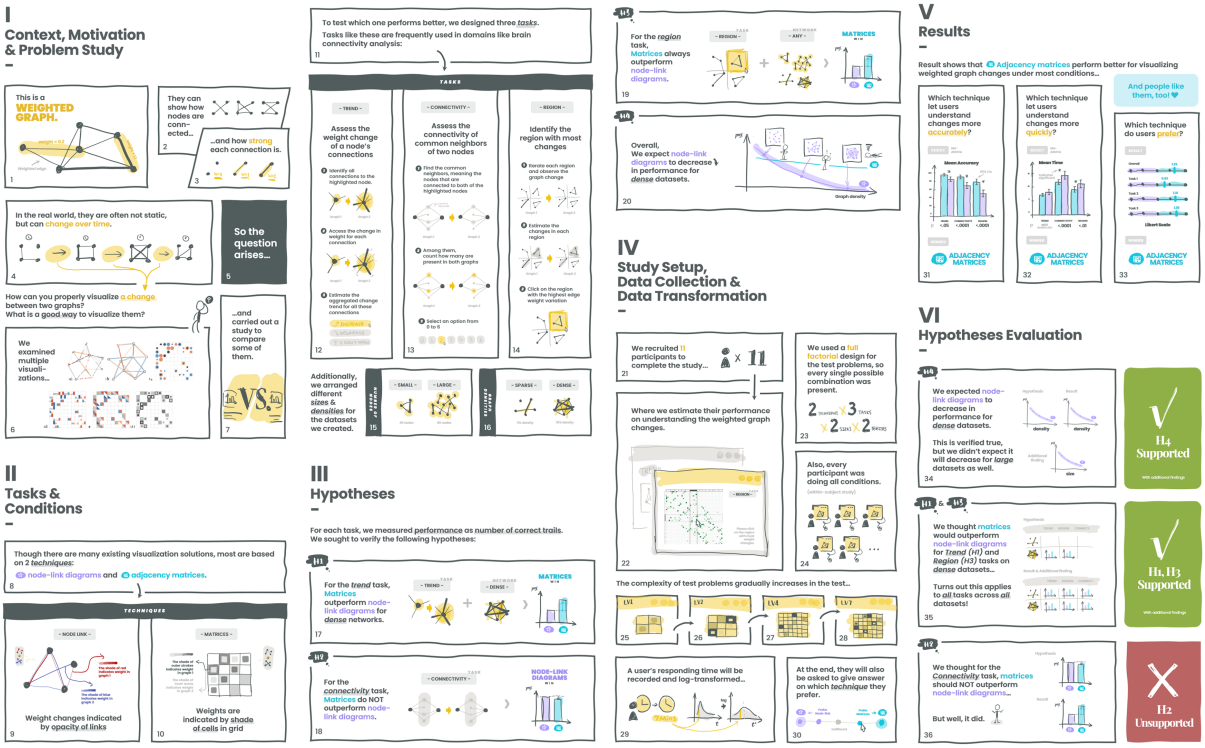


Figure 9.4: Example of a data comic illustrating a study that compared two visualization techniques for weighted networks. Created by one of the coauthors through an iterative process, based on information provided in the original paper [Alper et al., 2013]. Each panel is labeled with a number in the bottom left corner, which we reference in subsection 7.3.3.

and process. As an alternative solution to complement the oversimplification of graphical abstract [Hullman and Bach, 2018], we focus on *RQ5) How to use data comics for promoting real-world communication challenges, e.g., reporting quantitative user study and data analysis*. We provide examples of data comics for study reports (e.g., Figure 9.4) along with a framework of guidance for creating data comics for communicating common stages of HCI studies. The evaluation in collaborative design sessions shows our design solutions can help people to create compelling data comics for their study reports.

Chapter 8: Interactive Data Comics Interactions are commonly used in data visualizations that enable users to explore different data sets, change aspects, acquire information on demand. In data-driven stories, interactions could promote active exploration, self-reflection, so that to have deeper connections with audiences. However, data comics were mostly static by the time we investigated interactive data comics. This chapter seeks to explore *RQ6) What could interactions add to data comics* and *RQ7) How to create interactive data comics*. We provide a set of *operations* to support different narrative goals such as navigate, details on demand, change perspective, pause & reveal branch, and input data. To apply these operations in

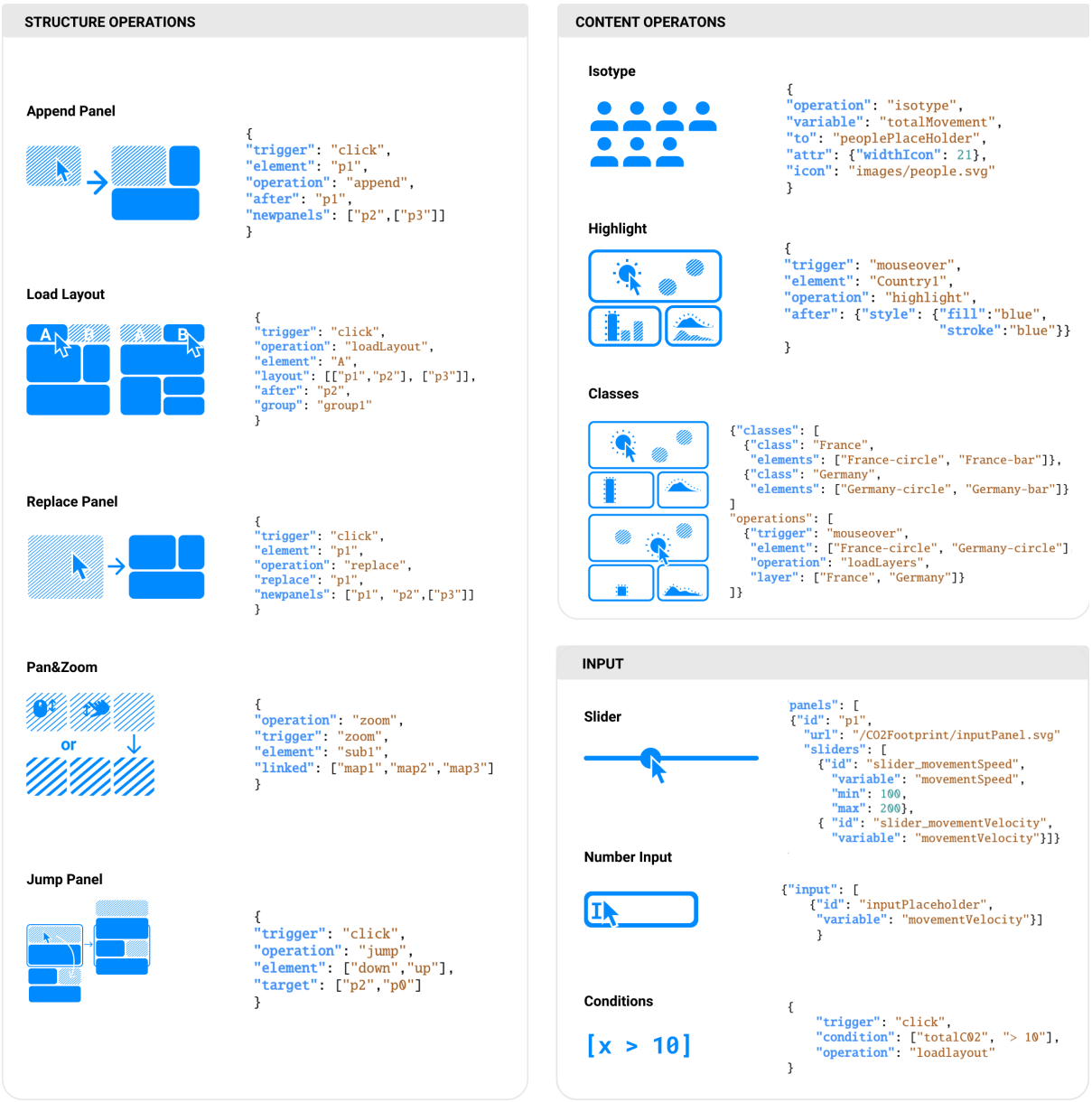


Figure 9.5: Example of interactive operations and corresponding code of ComicScript. Full descriptions detailed in section 8.6.

comics, we develop ComicScript for designers with light-weighted programming to build interactive data comics. Figure 9.5 demonstrates operations of three categories along with code of ComicScript of each one. We validate our authoring approach with professional designers, illustrators, and data comic enthusiasts, result show creators with little programming skills could create interactive data comics. ComicScript is our first step towards understanding interactive data comics, we are working on a graphical user interface to lower the effort of creation.

9.2 Limitation

Overall, this thesis explored the creation of data comics after the process of exploring raw data. The data comics creation process we investigated did not include the process of data analysis, so this thesis does not provide tools or method supports for creators to distill insights from raw data. In the current practice, the elements in the data comics are created by multiple approaches such as design tools, visualization creation tools, pen, and paper. We did not provide tools for creating data comics or elements therein. Also, the creation methods and pedagogy presented in this thesis could not improve the creators' skills in crafting individual components in data comics, for instance, building a story, illustration drawing, or creating data visualizations. The current supports do not cover all the components involved in data comics, for example, characters, metaphors, and humor are commonly found in traditional comics, yet we have not been explored their usage in data comics. These limitations require further research to address.

9.3 Agenda for Future Work

Iterating and Evaluating Data Comic Creation Model

The Data Comic Creation Model is currently a conceptual model inspired by a working model of user-experience design and combined with the knowledge gained during this PhD. The model potentially provokes new methods, materials, pedagogy, and tools of creating visual data stories yet have not been systematically evaluated. To validate its usage, and bridge the gap between the theoretical framework to actual practice, Data Comic Creation Model requires evaluation about its effectiveness in helping creators with creating data comics. For example, a series of workshops could be conducted, recruiting participants with different professionals working in teams to create data comics. In the workshop, use Data Comic Creation Model as the structured guidance, introducing goals, activities, and guidelines of each layer, eventually, interview participants to reflect on challenges of each step, receive feedback and improve the model, modify and crowd-source guidelines from participants. Once the model is validated and improved in the workshops, it can be evaluated as a theory-based pedagogy, e.g., providing a data-driven visual story creation MOOC for wider students, the model will be actively improved according to the outcomes and comments from the enrolled students.

A study of visual styles in data comics

The visual style is on the surface layer of the Data Comic Creation Model, playing a significant role in data comics, an appealing visual representation is the most mentioned factor that attracted pedestrians in the wild (Chapter 3). This thesis does not put the main research

focus on the elements of the *Graphic Appearance* layer. For example, would different visual styles affect how the audience feels about the presented data? Would sketchy style impair the trustworthiness of the data, or elevate its persuasive power? What are optimal graphic design choices for the different target audiences? Answering these questions can help creators to make better design decisions, and tools to provide style guidance or algorithms for automatic solutions.

Providing automatic or semi-automatic tool support for data comic creation

A structured framework and design guidance can streamline the process of data comic creation, yet manual creation is still time-consuming and laborious. Semi-automatic and automatic creating approaches could largely reduce time and effort. Although we are still far away from generating data comics for designated cases by clicking one button, yet by taking a step back, we can provide reusable templates and set parameters to reproduce data comics. However, reusing templates by changing data input requires a strict match of data structure, visualization techniques, and context. More flexible use of templates can be provided with adjustable modules and layouts, e.g., providing alternative visualization techniques and design patterns. Other ways of automatizing data comics include generating comics from written texts; transforming sketchy data comics into polished data comics with alternative styles; converting from other data-driven storytelling genres to data comics.

Extending the scenarios and applications of data comics

A growing number of practitioners and researchers are contributing new works of data comics and creation supports. Besides the scenarios data comics are normally applied, i.e., reporting, informing, and conveying information, data comics are not seen used as an approach more than presentation. Similar to storyboards being used in film direction and understanding users and products in UX and product design, we see the potential of using data comics as an approach to speculate ideas, evoking discussion, crowd-sourcing comments from audiences, and fostering active exploring. The Data Comic Creation Model in Section 3.1 is primarily inspired by the human-centered design approach, which sees designing visual data-driven stories as designing goal-oriented products, could data comics be data-centered or idea-centered?

Bibliography

Parallel coordinates. Available at: https://en.wikipedia.org/wiki/Parallel_coordinates. [online: last accessed January 14, 2022].

Flourish. Available at: https://flourish.studio/. [online: last accessed January 14, 2022].

Clip studio paint. Available at: https://www.celsys.co.jp/en/, 2013. [online: last accessed January 14, 2022].

Wikipedia panel (comics). https://en.wikipedia.org/wiki/Panel_(comics), 2017. [online: last accessed January 14, 2022].

Olympics athlete ages in summer 1908. Available at: http://www.zingchart.com/gallery/chart/#!vertical-boxplot-html-tooltip, 2019. [online: last accessed January 14, 2022].

Box and whisker plot. Available at: https://www.brainingcamp.com/lessons/box-and-whisker-plots/lesson.php, 2019. [online: last accessed January 14, 2022].

Car parameters. Available at: https://stackoverflow.com/questions/19213961/parallel-coordinates-program-written-with-processing-cant-show-anything-in-mac, 2019. [online: last accessed January 14, 2022].

Data science cheatsheets from datacamp. Available at: https://www.datacamp.com/community/data-science-cheatsheets, 2019. [online: last accessed January 14, 2022].

Fisher's irish data. Available at: https://en.wikipedia.org/wiki/File:ParCorFisherIris.png, 2019. [online: last accessed January 14, 2022].

Oxford dictionaries. Available at: https://en.oxforddictionaries.com/definition, 2019. [online: last accessed January 14, 2022].

Distill. Available at: https://distill.pub/, 2020. [online: last accessed January 14, 2022].

Experimental design and analysis of variance. Available at: https://yieldingresults.org/modules/module-2/, 2020. [online: last accessed January 14, 2022].

Seeing theory. Available at: https://seeing-theory.brown.edu/, 2020. [online: last accessed January 14, 2022].

Publication Manual of the American Psychological Association, Seventh Edition. American Psychological Association, 2020. URL https://apastyle.apa.org/products/publication-manual-7th-edition/?tab=5.

medium.com. Available at: https://medium.com/, 2020. [online: last accessed January 14, 2022].

researchblogging.org. Available at: http://researchblogging.org/, 2020. [online: last accessed January 14, 2022].

Comic life. Available at: http://plasq.com/, 2021. [online: last accessed January 14, 2022].

Gregor Aisch. Seven features you'll want in your next charting tool. *vis4. net. https://www. vis4. net/blog/2015/03/seven-features-youll-wanfin-your-next-charfing-tool*, 2015.

Paul A Aleixo and Krystina Sumner. Memory for biopsychology material presented in comic book format. *Journal of graphic novels & comics*, 8(1):79–88, 2017. ISSN 2150-4857.

Danielle B. Allen. *Storytelling with data: A data visualization guide for business professionals*, volume 72. 2019. doi: https://doi.org/10.1111/peps.12333. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/peps.12333.

Basak Alper, Benjamin Bach, Nathalie Henry Riche, Tobias Isenberg, and Jean-Daniel Fekete. Weighted graph comparison techniques for brain connectivity analysis. In *Proceedings of the 2013 CHI Conference on Human Factors in Computing Systems*, CHI '13, page 483–492, New York, NY, USA, 2013. Association for Computing Machinery. ISBN 9781450318990. doi: 10.1145/2470654.2470724. URL https://doi.org/10.1145/2470654.2470724.

Sara Varela Amaral, Teresa Forte, João Ramalho-Santos, and M Teresa Girao da Cruz. I want more and better cells!–an outreach project about stem cells and its impact on the general population. *PloS one*, 10(7), 2015.

Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Christophe Hurter, and Pourang Irani. 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*, CHI '15, page 1459–1468, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450331456. doi: 10.1145/2702123.2702431. URL https://doi.org/10.1145/2702123.2702431.

Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Andres Monroy-Hernandez, and Pourang Irani. Authoring data-driven videos with dataclips. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):501–510, 2017. doi: 10.1109/TVCG.2016.2598647.

E. W. Anderson, K. C. Potter, L. E. Matzen, J. F. Shepherd, G. A. Preston, and C. T. Silva. A user study of visualization effectiveness using eeg and cognitive load. volume 30, pages 791–800, 2011. doi: https://doi.org/10.1111/j.1467-8659.2011.01928.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8659.2011.01928.x.

Daniel Andrews, Chris Baber, Sergey Efremov, and Mikhail Komarov. Creating and using interactive narratives: Reading and writing branching comics. In *Proceedings of the 2012 CHI Conference on Human Factors in Computing Systems*, CHI '12, page 1703–1712, New York, NY, USA, 2012. Association for Computing Machinery. ISBN 9781450310154. doi: 10.1145/2207676.2208298. URL https://doi.org/10.1145/2207676.2208298.

Benjamin Bach, Emmanuel Pietriga, and Jean-Daniel Fekete. Graphdiaries: Animated transitions andtemporal navigation for dynamic networks. *IEEE Transactions on Visualization and Computer Graphics*, 20(5):740–754, 2014a. doi: 10.1109/TVCG.2013.254.

Benjamin Bach, Emmanuel Pietriga, and Jean-Daniel Fekete. Visualizing dynamic networks with matrix cubes. In *Proceedings of the 2014 CHI Conference on Human Factors in Computing Systems*, CHI '14, page 877–886, New York, NY, USA, 2014b. Association for Computing Machinery. ISBN 9781450324731. doi: 10.1145/2556288.2557010. URL https://doi.org/10.1145/2556288.2557010.

Benjamin Bach, Nathalie Henry-Riche, Tim Dwyer, Tara Madhyastha, J-D Fekete, and Thomas Grabowski. Small multipiles: Piling time to explore temporal patterns in dynamic networks. In *Computer Graphics Forum*, volume 34, pages 31–40, 2015.

Benjamin Bach, Natalie Kerracher, Kyle Wm. Hall, Sheelagh Carpendale, Jessie Kennedy, and Nathalie Henry Riche. Telling stories about dynamic networks with graph comics. In

Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, page 3670–3682, New York, NY, USA, 2016a. Association for Computing Machinery. ISBN 9781450333627. doi: 10.1145/2858036.2858387. URL https://doi.org/10.1145/2858036.2858387.

Benjamin Bach, Conglei Shi, Nicolas Heulot, Tara Madhyastha, Tom Grabowski, and Pierre Dragicevic. Time curves: Folding time to visualize patterns of temporal evolution in data. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):559–568, 2016b. doi: 10.1109/TVCG.2015.2467851.

Benjamin Bach, Nathalie Henry Riche, Sheelagh Carpendale, and Hanspeter Pfister. The emerging genre of data comics. *IEEE Computer Graphics and Applications*, 37(3):6–13, 2017a. doi: 10.1109/MCG.2017.33.

Benjamin Bach, Nathalie Henry Riche, Christophe Hurter, Kim Marriott, and Tim Dwyer. Towards unambiguous edge bundling: Investigating confluent drawings for network visualization. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):541–550, 2017b. doi: 10.1109/TVCG.2016.2598958.

Benjamin Bach, Zezhong Wang, Nathalie Henry Riche, Matteo Farinella, Dave Murray-Rust, Sheelagh Carpendale, and Hanspeter Pfister. Data comics. Available at: https://datacomics.github.io/, 2017-2020. [online: last accessed January 14, 2022].

Benjamin Bach, Ronell Sicat, Johanna Beyer, Maxime Cordeil, and Hanspeter Pfister. The hologram in my hand: How effective is interactive exploration of 3d visualizations in immersive tangible augmented reality? *IEEE Transactions on Visualization and Computer Graphics*, 24 (1):457–467, 2018a. doi: 10.1109/TVCG.2017.2745941.

Benjamin Bach, Moritz Stefaner, Jeremy Boy, Steven Drucker, Lyn Bartram, Jo Wood, Paolo Ciuccarelli, Yuri Engelhardt, Ulrike Koeppen, and Barbara Tversky. Narrative design patterns for data-driven storytelling. In *Data-Driven Storytelling*, pages 125–152. AK Peters/CRC Press, 2018b.

Benjamin Bach, Zezhong Wang, Matteo Farinella, Dave Murray-Rust, and Nathalie Henry Riche. Design patterns for data comics. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 1–12, New York, NY, USA, 2018c. Association for Computing Machinery. ISBN 9781450356206. URL https://doi.org/10.1145/3173574.3173612.

Sriram Karthik Badam, Zhicheng Liu, and Niklas Elmqvist. Elastic documents: Coupling text and tables through contextual visualizations for enhanced document reading. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):661–671, 2019. doi: 10.1109/TVCG.2018.2865119.

Alan D. Baddeley. *Human memory : theory and practice / Alan Baddeley.* Psychology Press, Hove, revised edition. edition, 1997. ISBN 0863774318.

Anouk Barberousse. What is the use of diagrams in theoretical modeling? *Science in context*, 26(2):345–362, 2013. ISSN 0269-8897.

Ashok R. Basawapatna, Kyu Han Koh, and Alexander Repenning. Using scalable game design to teach computer science from middle school to graduate school. In *Proceedings of the Fifteenth Annual Conference on Innovation and Technology in Computer Science Education*, ITiCSE '10, page 224–228, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 9781605588209. doi: 10.1145/1822090.1822154. URL https://doi.org/10.1145/1822090.1822154.

Scott Bateman, Regan L. Mandryk, Carl Gutwin, Aaron Genest, David McDine, and Christopher Brooks. Useful junk? the effects of visual embellishment on comprehension and memorability of charts. In *Proceedings of the 2010 CHI Conference on Human Factors in Computing Systems*, CHI '10, page 2573–2582, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 9781605589299. doi: 10.1145/1753326.1753716. URL https://doi.org/10.1145/1753326.1753716.

Colin B Begg. Publication bias. *The handbook of research synthesis*, 25:299–409, 1994.

Michael Behrisch, Benjamin Bach, Nathalie Henry Riche, Tobias Schreck, and Jean-Daniel Fekete. Matrix reordering methods for table and network visualization. volume 35, pages 693–716, 2016. doi: https://doi.org/10.1111/cgf.12935. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.12935.

Stephanie Bell. Project-based learning for the 21st century: Skills for the future. *The Clearing house*, 83(2):39–43, 2010. ISSN 0009-8655.

Arnd Bernaerts. Booklet on naval war changes climate: A fascinating theory on the impact of naval warfare on climate. 2006.

Jacques Bertin, William J Berg, and Howard Wainer. *Semiology of graphics: diagrams, networks, maps*, volume 1. University of Wisconsin press Madison, 1983.

Rebecca Bertram and Radostina Primova. *Energy Atlas 2018: Figures and Facts about Renewables in Europe*. Heinrich Böll Foundation, 2018.

Ruth Beyth-Marom, Fiona Fidler, and Geoff Cumming. Statistical cognition: Towards evidence-base practice in statistics and statistics education. *Statistics Education Research Journal*, 7(2), 2008.

Alex Bigelow, Steven Drucker, Danyel Fisher, and Miriah Meyer. Iterating between tools to create and edit visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):481–490, 2017. doi: 10.1109/TVCG.2016.2598609.

John B Black and Gordon H Bower. Episodes as chunks in narrative memory. *Journal of verbal learning and verbal behavior*, 18(3):309–318, 1979.

Michelle A. Borkin, Azalea A. Vo, Zoya Bylinskii, Phillip Isola, Shashank Sunkavalli, Aude Oliva, and Hanspeter Pfister. What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2306–2315, 2013. doi: 10.1109/TVCG.2013.234.

Katy Börner, Adam Maltese, Russell Nelson Balliet, and Joe Heimlich. Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Visualization*, 15(3):198–213, 2016.

M Bostock. D3.js-data-driven documents. Available at: `https://d3js.org/`, 2019. [online: last accessed January 14, 2022].

Mike Bostock. Les misérables co-occurrence. Available at: `https://bost.ocks.org/mike/miserables/`, 2012. [online: last accessed January 14, 2022].

Jeremy Boy, Ronald A. Rensink, Enrico Bertini, and Jean-Daniel Fekete. A principled way of assessing visualization literacy. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1963–1972, 2014. doi: 10.1109/TVCG.2014.2346984.

Jeremy Boy, Louis Eveillard, Françoise Detienne, and Jean-Daniel Fekete. Suggested interactivity: Seeking perceived affordances for information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):639–648, 2016. doi: 10.1109/TVCG.2015.2467201.

Valerie J Bristor and Suzanne V Drake. Linking the language arts and content areas through visual technology. *THE Journal (Technological Horizons In Education)*, 22(2):74, 1994.

Kieron Brown. Musical sequences in comics. *The Comics Grid: Journal of Comics Scholarship*, 3(1), 2013.

Susie Cagle. Humans have made 8.3bn tons of plastic since 1950. Available at: https://www.theguardian.com/us-news/2019/jun/23/all-the-plastic-ever-made-study-comic, 2019. [online: last accessed January 14, 2022].

Alberto Cairo. *The functional art : an introduction to information graphics and visualization / Alberto Cairo.* New Riders, Berkely, Calif, 2013 - 2013. ISBN 9780321834737.

Alberto Cairo. *The truthful art : data, charts, and maps for communication / Alberto Cairo.* New Riders, San Francisco, California, 2016 - 2016. ISBN 9780133440539.

Joshua Caldwell. Information comics: An overview. In *2012 IEEE International Professional Communication Conference*, pages 1–7, 2012. doi: 10.1109/IPCC.2012.6408645.

Michael P. J Camilleri and Christopher K. I Williams. The extended dawid-skene model: Fusing information from multiple data schemas. *Machine Learning and Knowledge Discovery in Databases*, 1167:121–136, 2020. ISSN 1865-0929.

Ying Cao, Antoni B. Chan, and Rynson W. H. Lau. Automatic stylistic manga layout. *ACM Trans. Graph.*, 31(6), nov 2012. ISSN 0730-0301. doi: 10.1145/2366145.2366160. URL https://doi.org/10.1145/2366145.2366160.

Stuart K. Card. *Readings in information visualization : using vision to think.* The Morgan Kaufmann series in interactive technologies. Morgan Kaufmann Publishers, San Francisco, Calif, 1999. ISBN 1558605339.

David Carr. Guidelines for designing information visualization applications. 1999.

Iain Chalmers and Paul Glasziou. Avoidable waste in the production and reporting of research evidence. *The Lancet*, 374(9683):86–89, 2009. ISSN 0140-6736. doi: https://doi.org/10.1016/S0140-6736(09)60329-9. URL https://www.sciencedirect.com/science/article/pii/S0140673609603299.

Jorge Cham. Phd comics. Available at: http://phdcomics.com/. [online: last accessed January 14, 2022].

Jorge Cham. The science gap. Available at: https://tedx.ucla.edu/talks/jorge_cham_the_science_gap/, 2012. [online: last accessed January 14, 2022].

Chan, Long Ting. Leveraging the proteus effect to motivate emotional support in a serious game for mental health, 2019. URL http://hdl.handle.net/10012/15199.

Paul Chandler and John Sweller. Cognitive load theory and the format of instruction. *Cognition and instruction*, 8(4):293–332, 1991.

Bay-Wei Chang, Jock Mackinlay, and Polle T Zellweger. Fluidly revealing information in fluid documents. In *Proc. Smart Graphics AAAI Spring Symposium*, 2000.

Chunlei Chang, Benjamin Bach, Tim Dwyer, and Kim Marriott. Evaluating perceptually complementary views for network exploration tasks. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 1397–1407, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450346559. doi: 10.1145/3025453.3026024. URL https://doi.org/10.1145/3025453.3026024.

Bingxin Chen, Rebecca Jablonsky, Jack Benjamin Margines, Raunaq Gupta, and Shailie Thakkar. Comic circuit: an online community for the creation and consumption of news comics. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 2561–2566. 2013.

Lushi Chen, Walid Magdy, Heather Whalley, and Maria Klara Wolters. Examining the role of mood patterns in predicting self-reported depressive symptoms. page 164–173, 2020. doi: 10.1145/3394231.3397906. URL https://doi.org/10.1145/3394231.3397906.

Fanny Chevalier, Romain Vuillemot, and Guia Gali. Using concrete scales: A practical framework for effective visual depiction of complex measures. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2426–2435, 2013. doi: 10.1109/TVCG.2013.210.

Fanny Chevalier, Pierre Dragicevic, and Steven Franconeri. The not-so-staggering effect of staggered animated transitions on visual tracking. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2241–2250, 2014. doi: 10.1109/TVCG.2014.2346424.

Fanny Chevalier, Nathalie Henry Riche, Catherine Plaisant, Amira Chalbi, and Christophe Hurter. Animations 25 years later: New roles and opportunities. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '16, page 280–287, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450341318. doi: 10.1145/2909132.2909255. URL https://doi.org/10.1145/2909132.2909255.

Mike Cisneros. Is that true? Available at: https://public.tableau.com/profile/mikevizneros#!/vizhome/IsThatRight/IsThatTrue, 2011. [online: last accessed January 14, 2022].

James M Clark and Allan Paivio. Dual coding theory and education. *Educational psychology review*, 3(3):149–210, 1991. doi: https://doi.org/10.1007/BF01320076.

Neil Cohn. Navigating comics: an empirical and theoretical approach to strategies of reading comic page layouts. *Frontiers in psychology*, 4:186–186, 2013a. ISSN 1664-1078.

Neil Cohn. *Visual Language of Comics: Introduction to the Structure and Cognition of Sequential Images*. Bloomsbury advances in semiotics. Bloomsbury Publishing, New York, 2013b. ISBN 1441170545.

Neil Cohn. The architecture of visual narrative comprehension: the interaction of narrative structure and page layout in understanding comics. *Frontiers in Psychology*, 5, 2014. ISSN 1664-1078. doi: 10.3389/fpsyg.2014.00680. URL http://journal.frontiersin.org/article/10.3389/fpsyg.2014.00680/abstract.

Matthew Conlen and Jeffrey Heer. Idyll: A markup language for authoring and publishing interactive articles on the web. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, UIST '18, page 977–989, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450359481. doi: 10.1145/3242587.3242600. URL https://doi.org/10.1145/3242587.3242600.

Matthew Conlen, Megan Vo, Alan Tan, and Jeffrey Heer. *Idyll Studio: A Structured Editor for Authoring Interactive Data-Driven Articles*, page 1–12. Association for Computing Machinery, New York, NY, USA, 2021. ISBN 9781450386357. URL https://doi.org/10.1145/3472749.3474731.

Kristin A Cook and James J Thomas. Illuminating the path: The research and development agenda for visual analytics. Technical report, Pacific Northwest National Lab.(PNNL), Richland, WA (United States), 2005.

Michael Correll and Michael Gleicher. Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2142–2151, 2014. doi: 10.1109/TVCG.2014.2346298.

Design Council. Eleven lessons: Managing design in eleven global companies-desk research report. *Design Council*, 2007.

Design Council. The double diamond. Available at: https://www.designcouncil.org.uk/news-opinion/design-methods-step-1-discover, 2015. [online: last accessed January 14, 2022].

Amanda Cox. How editing and design changes news graphics. In *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages xiii–xiii. IEEE, 2011. ISBN 1467300152.

Nigel Cross. *Engineering design methods*. Wiley, Chichester, 1989.

Pedro Cruz and Penousal Machado. Generative storytelling for information visualization. *IEEE Computer Graphics and Applications*, 31(2):80–85, 2011. doi: 10.1109/MCG.2011.23.

Paul Fisher Davies. *Coda: Metaphor, Magic and Making Meanings*. Springer, Cham, 2019. ISBN 978-3-030-29722-0. doi: 10.1007/978-3-030-29722-0_8. URL https://doi.org/10.1007/978-3-030-29722-0_8.

C. Delp and J. Jones. Communicating information to patients: the use of cartoon illustrations to improve comprehension of instructions. *Academic Emergency Medicine: Official Journal of the Society for Academic Emergency Medicine*, 3(3):264–270, March 1996. ISSN 1069-6563.

Nadine Desrochers, Adèle Paul-Hus, Stefanie Haustein, Rodrigo Costas, Philippe Mongeon, Anabel Quan-Haase, Timothy D Bowman, Jen Pecoskie, Andrew Tsou, and Vincent Larivière. Authorship, citations, acknowledgments and visibility in social media: Symbolic capital in the multifaceted reward system of science. *Social science information*, 57(2):223–248, 2018.

Judy Diamond, Julia McQuillan, Amy N. Spiegel, Patricia Wonch Hill, Rebecca Smith, John West, and Charles Wood. Viruses, Vaccines and the Public. *Museums & Social Issues*, 11 (1):9–16, January 2016. ISSN 1559-6893. doi: 10.1080/15596893.2016.1131099. URL http://dx.doi.org/10.1080/15596893.2016.1131099.

Thomas N Dorsel and Gary W Cundiff. The cheat-sheet: Efficient coding device or indispensable crutch? *The Journal of Experimental Education*, 48(1):39–42, 1979.

Pierre Dragicevic. Fair statistical communication in hci. In Judy Robertson and Maurits Kaptein, editors, *Modern Statistical Methods for HCI*, pages 291–330. Springer, Cham, 2016. ISBN 978-3-319-26633-6. doi: 10.1007/978-3-319-26633-6_13. URL https://doi.org/10.1007/978-3-319-26633-6_13.

Pierre Dragicevic, Yvonne Jansen, Abhraneel Sarma, Matthew Kay, and Fanny Chevalier. Increasing the transparency of research papers with explorable multiverse analyses. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, page 1–15, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450359702. URL https://doi.org/10.1145/3290605.3300295.

Pierre Dragicevic, Yvonne Jansen, and Andrew Vande Moere. *Data Physicalization*. Springer, Cham, 2020. ISBN 978-3-319-27648-9. doi: 10.1007/978-3-319-27648-9_94-1. URL https://doi.org/10.1007/978-3-319-27648-9_94-1.

Nancy. Duarte. *Resonate present visual stories that transform audiences / Nancy Duarte.* Wiley, Hoboken, NJ, 1st edition edition, 2010. ISBN 1-118-00893-6.

Abigail Durrant, Duncan Rowland, David S. Kirk, Steve Benford, Joel E. Fischer, and Derek McAuley. Automics: Souvenir generating photoware for theme parks. In *Proceedings of the 2011 CHI Conference on Human Factors in Computing Systems*, CHI '11, page 1767–1776, New York, NY, USA, 2011. Association for Computing Machinery. ISBN 9781450302289. doi: 10.1145/1978942.1979199. URL https://doi.org/10.1145/1978942.1979199.

Ryan Eccles, Thomas Kapler, Robert Harper, and William Wright. Stories in geotime. *Information Visualization*, 7(1):3–17, 2008.

The Economist. The beef with beef: Treating beef like coal would make a big dent in greenhouse-gas emissions. Available at:https://www.economist.com/graphic-detail/2021/10/02/treating-beef-like-coal-would-make-a-big-dent-in-greenhouse-gas-emissions. [online: last accessed January 14, 2022].

Alexander Eiselmayer, Chat Wacharamanotham, Michel Beaudouin-Lafon, and Wendy E. Mackay. Touchstone2: An interactive environment for exploring trade-offs in hci experiment design. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 217:1–217:11, New York, NY, USA, 2019. ACM. ISBN 978-1-4503-5970-2. doi: 10.1145/3290605.3300447. URL http://doi.acm.org/10.1145/3290605.3300447.

Will Eisner. *Graphic storytelling and visual narrative*. WW Norton & Company, 2008.

Niklas Elmqvist, Andrew Vande Moere, Hans-Christian Jetter, Daniel Cernea, Harald Reiterer, and TJ Jankun-Kelly. Fluid interaction for information visualization. *Information Visualization*, 10(4):327–340, 2011.

Phillip Endicott, Simon Y.W. Ho, and Chris Stringer. Using genetic evidence to evaluate four palaeoanthropological hypotheses for the timing of neanderthal and modern human origins. *Journal of Human Evolution*, 59(1):87–95, 2010. ISSN 0047-2484. doi: https://doi.org/10.1016/j.jhevol.2010.04.005. URL https://www.sciencedirect.com/science/article/pii/S0047248410000710.

Brigitte Erbe. Reducing test anxiety while increasing learning: The cheat sheet. *College teaching*, 55(3):96–98, 2007.

Stephanie D. H. Evergreen and Gavin McMahon. Data visualization chart chooser cards. Available at: http://chartchoosercards.com/, 2019. [online: last accessed January 14, 2022].

Matteo Farinella. The potential of comics in science communication. *Journal of science communication*, 17(1):Y01, 2018.

Anna Feigenbaum and Aria Alamalhodaei. *The Data Storytelling Workbook*. Routledge, 2020.

Ferdio. Datavizproject. Available at: https://datavizproject.com, 2019. [online: last accessed January 14, 2022].

Stephen Few. Designing effective tables and graphs, 2012.

Tom Foulsham, Dean Wybrow, and Neil Cohn. Reading without words: Eye movements in the comprehension of comic strips. *Applied Cognitive Psychology*, 30(4):566–579, 2016. doi: https://doi.org/10.1002/acp.3229. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/acp.3229.

Steve Franconeri. Which visualization? a quick reference. Available at: http://experception.net, 2018. [online: last accessed January 14, 2022].

Gustav Freytag. *Technique of the drama: An exposition of dramatic composition and art*. S. Griggs, 1895.

Johanna Fulda, Matthew Brehmer, and Tamara Munzner. Timelinecurator: Interactive authoring of visual timelines from unstructured text. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):300–309, 2016. doi: 10.1109/TVCG.2015.2467531.

Tong Gao, Jessica R. Hullman, Eytan Adar, Brent Hecht, and Nicholas Diakopoulos. Newsviews: An automated pipeline for creating custom geovisualizations for news. In *Proceedings of the 2014 CHI Conference on Human Factors in Computing Systems*, CHI '14, page 3005–3014, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450324731. doi: 10.1145/2556288.2557228. URL https://doi.org/10.1145/2556288.2557228.

Jesse James Garrett. *Elements of user experience, the: user-centered design for the web and beyond*. Pearson Education, 2010.

Nahum Gershon and Ward Page. What storytelling can do for information visualization. *Commun. ACM*, 44(8):31–37, aug 2001. ISSN 0001-0782. doi: 10.1145/381641.381653. URL https://doi.org/10.1145/381641.381653.

Larry Gonick and Woollcott Smith. *The cartoon guide to statistics*. Collins, 1993.

Daniel Merlin Goodbrey. From comic to hypercomic. In *cultural excavation and formal expression in the graphic novel*, pages 291–302. Brill, 2013.

David Graeber. *The utopia of rules: On technology, stupidity, and the secret joys of bureaucracy*. Melville House, 2015.

Gramener. Comicgen. Available at: https://gramener.com/comicgen/#, 2019. [online: last accessed January 14, 2022].

Jonathan Gray, Lucy Chambers, and Liliana Bounegru. *The data journalism handbook: How journalists can use data to improve the news*. O'Reilly Media, Inc, 2012.

Melanie C. Green, Jeffrey J. Strange, and Timothy C. Brock. *Narrative Impact: Social and Cognitive Foundations*. Taylor & Francis, January 2003. ISBN 978-1-135-67328-4.

Thierry Groensteen. *The System of Comics*. Univ. Press of Mississippi, 2007. ISBN 978-1-60473-693-9.

Thierry Groensteen. *Comics and Narration*. Univ. Press of Mississippi, 2013.

Tovi Grossman, Fanny Chevalier, and Rubaiat Habib Kazi. Your paper is dead! bringing life to research articles with animated figures. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, page 461–475, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450331463. doi: 10.1145/2702613.2732501. URL https://doi.org/10.1145/2702613.2732501.

CONSORT Group. Consort: Consolidated standards of reporting trials. Available at: http://www.consort-statement.org/, 2010. [online: last accessed January 14, 2022].

Steve Haroz and David Whitney. How capacity limits of attention influence information visualization effectiveness. *IEEE Transactions on Visualization and Computer Graphics*, 18 (12):2402–2410, 2012. doi: 10.1109/TVCG.2012.233.

Steve Haroz, Robert Kosara, and Steven L. Franconeri. The connected scatterplot for presenting paired time series. *IEEE Transactions on Visualization and Computer Graphics*, 22(9): 2174–2186, 2016. doi: 10.1109/TVCG.2015.2502587.

Shiqing He and Eytan Adar. Vizitcards: A card-based toolkit for infovis design education. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):561–570, 2017. doi: 10.1109/TVCG.2016.2599338.

Conor Healy. data-to-viz.com. Available at: `https://www.data-to-viz.com/`, 2019. [online: last accessed January 14, 2022].

Marti A Hearst. Active learning assignments for student acquisition of design principles. In *Pedagogy Data Visualization, IEEE VIS Workshop*, 2016.

Jeffrey Heer and George Robertson. Animated transitions in statistical data graphics. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1240–1247, 2007. doi: 10.1109/TVCG.2007.70539.

Jeffrey Heer and Ben Shneiderman. Interactive dynamics for visual analysis. *Commun. ACM*, 55(4):45–54, apr 2012. ISSN 0001-0782. doi: 10.1145/2133806.2133821. URL `https://doi.org/10.1145/2133806.2133821`.

Jeffrey Heer, Nicholas Kong, and Maneesh Agrawala. Sizing the horizon: The effects of chart size and layering on the graphical perception of time series visualizations. In *Proceedings of the 2009 CHI Conference on Human Factors in Computing Systems*, CHI '09, page 1303–1312, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605582467. doi: 10.1145/1518701.1518897. URL `https://doi.org/10.1145/1518701.1518897`.

Nathalie Henry and Jean-daniel Fekete. Matrixexplorer: a dual-representation system to explore social networks. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):677–684, 2006. doi: 10.1109/TVCG.2006.160.

Sarada Herke. Graph theory faqs: 03. isomorphism using adjacency matrix. Available at: `https://www.youtube.com/watch?v=UCle3Smvh1s&t=27s`, 2019. [online: last accessed January 14, 2022].

Jacob Hilton, Nick Cammarata, Shan Carter, Gabriel Goh, and Chris Olah. Understanding rl vision. *Distill*, 5(11):e29, 2020.

Uta Hinrichs, Holly Schmidt, and Sheelagh Carpendale. Emdialog: Bringing information visualization into the museum. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1181–1188, 2008. doi: 10.1109/TVCG.2008.127.

Wei-Chin Ho, Yoshikazu Ohya, and Jianzhi Zhang. Testing the neutral hypothesis of phenotypic evolution. *Proceedings of the National Academy of Sciences*, 114(46):12219–12224, 2017.

Fred Hohman, Matthew Conlen, Jeffrey Heer, and Duen Horng Polo Chau. Communicating with interactive articles. *Distill*, 5(9), 2020.

Kasper Hornbæk, Søren S. Sander, Javier Andrés Bargas-Avila, and Jakob Grue Simonsen. Is once enough? on the extent and content of replications in human-computer interaction. In *Proceedings of the 2014 CHI Conference on Human Factors in Computing Systems*, CHI '14, page 3523–3532, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450324731. doi: 10.1145/2556288.2557004. URL https://doi.org/10.1145/2556288.2557004.

Jay Hosler and KB Boomer. Are comic books an effective way to engage nonmajors in learning and appreciating science? *CBE—Life Sciences Education*, 10(3):309–317, 2011.

I-Han Hsiao and Claudia López. Lessons learned from students' cheat sheets: Generic models for designing programming study guides. In *2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT)*, pages 209–211, 2016. doi: 10.1109/ICALT.2016.91.

Jessica Hullman and Benjamin Bach. Picturing science: Design patterns in graphical abstracts. In Peter Chapman, Gem Stapleton, Amirouche Moktefi, Sarah Perez-Kriz, and Francesco Bellucci, editors, *Diagrammatic Representation and Inference*, pages 183–200, Cham, 2018. Springer. ISBN 978-3-319-91376-6.

Jessica Hullman and Nick Diakopoulos. Visualization rhetoric: Framing effects in narrative visualization. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2231–2240, 2011. doi: 10.1109/TVCG.2011.255.

Jessica Hullman, Nicholas Diakopoulos, and Eytan Adar. Contextifier: Automatic generation of annotated stock visualizations. In *Proceedings of the 2013 CHI Conference on Human Factors in Computing Systems*, CHI '13, page 2707–2716, New York, NY, USA, 2013a. Association for Computing Machinery. ISBN 9781450318990. doi: 10.1145/2470654.2481374. URL https://doi.org/10.1145/2470654.2481374.

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

Jessica Hullman, Xiaoli Qiao, Michael Correll, Alex Kale, and Matthew Kay. In pursuit of error: A survey of uncertainty visualization evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):903–913, 2019. doi: 10.1109/TVCG.2018.2864889.

Samuel Huron, Sheelagh Carpendale, Jeremy Boy, and Jean-Daniel Fekete. Using viskit: A manual for running a constructive visualization workshop. In *Pedagogy of Data Visualization Workshop at IEEE VIS 2016*, 2016.

Samuel Huron, Pauline Gourlet, Uta Hinrichs, Trevor Hogan, and Yvonne Jansen. Let's get physical: Promoting data physicalization in workshop formats. In *Proceedings of the 2017 Conference on Designing Interactive Systems*, DIS '17, page 1409–1422, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450349222. doi: 10.1145/3064663.3064798. URL https://doi.org/10.1145/3064663.3064798.

Tim N. Höffler and Detlev Leutner. Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction*, 17(6):722–738, 2007. ISSN 0959-4752. doi: https://doi.org/10.1016/j.learninstruc.2007.09.013. URL https://www.sciencedirect.com/science/article/pii/S0959475207001077.

Noah Iliinsky. Visualization design thinking for effective data visualization. Available at: https://datascientistinsights.com/2012/12/05/design-thinking-for-effective-data-visualization/, 2019. [online: last accessed January 14, 2022].

Noah Iliinsky and Julie Steele. *Designing data visualizations: Representing informational Relationships*. O'Reilly Media, Inc, 2011.

US Infrastructure. Available at: http://aquadoc.typepad.com/waterwired/water_quantity/page/121/. Original website not available anymore., 2010. [online: last accessed January 14, 2022].

Laurent Itti and Christof Koch. Computational modelling of visual attention. *Nature reviews neuroscience*, 2(3):194–203, 2001.

Chris Johnson, Robert Moorhead, Tamara Munzner, Hanspeter Pfister, Penny Rheingans, and Terry S Yoo. Nih-nsf visualization research challenges report. Institute of Electrical and Electronics Engineers, 2005.

Gerald Everett Jones. *How to lie with charts*. LaPuerta Books and Media, 2011.

DaYe Kang, Tony Ho, Nicolai Marquardt, Bilge Mutlu, and Andrea Bianchi. Toonnote: Improving communication in computational notebooks using interactive data comics. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380966. doi: 10.1145/3411764.3445434. URL https://doi.org/10.1145/3411764.3445434.

Minjeong Kang. Measuring social media credibility: A study on a measure of blog credibility. *Institute for Public Relations*, pages 59–68, 2010.

Koray Karaca. Representing experimental procedures through diagrams at cern's large hadron collider: The communicatory value of diagrammatic representations in collaborative research. *Perspectives on science*, 25(2):177–203, 2017.

Helen Kennedy, Rosemary Lucy Hill, William Allen, and Andy Kirk. Engaging with (big) data visualizations: Factors that affect engagement and resulting new definitions of effectiveness. *First Monday*, 21(11), 2016.

Ethan Kerzner, Sarah Goodwin, Jason Dykes, Sara Jones, and Miriah Meyer. A framework for creative visualization-opportunities workshops. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):748–758, 2019. doi: 10.1109/TVCG.2018.2865241.

Carol Kilkenny, William J Browne, Innes C Cuthill, Michael Emerson, and Douglas G Altman. Improving bioscience research reporting: the arrive guidelines for reporting animal research. *PLoS biology*, 8(6), 2010.

Jiyoon Kim, Min Suk Chung, Hae Gwon Jang, and Beom Sun Chung. The use of educational comics in learning anatomy among multiple student groups. *Anatomical Sciences Education*, 10(1):79–86, 2017a. doi: https://doi.org/10.1002/ase.1619. URL https://anatomypubs.onlinelibrary.wiley.com/doi/abs/10.1002/ase.1619.

Nam Wook Kim, Eston Schweickart, Zhicheng Liu, Mira Dontcheva, Wilmot Li, Jovan Popovic, and Hanspeter Pfister. Data-driven guides: Supporting expressive design for information graphics. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):491–500, 2017b. doi: 10.1109/TVCG.2016.2598620.

Nam Wook Kim, Nathalie Henry Riche, Benjamin Bach, Guanpeng Xu, Matthew Brehmer, Ken Hinckley, Michel Pahud, Haijun Xia, Michael J. McGuffin, and Hanspeter Pfister. Datatoon: Drawing dynamic network comics with pen + touch interaction. page 1–12, 2019. URL https://doi.org/10.1145/3290605.3300335.

Yea-Seul Kim, Katharina Reinecke, and Jessica Hullman. Explaining the gap: Visualizing one's predictions improves recall and comprehension of data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 1375–1386, New York, NY, USA, 2017c. Association for Computing Machinery. ISBN 9781450346559. doi: 10.1145/3025453.3025592. URL https://doi.org/10.1145/3025453.3025592.

Younghoon Kim, Kanit Wongsuphasawat, Jessica Hullman, and Jeffrey Heer. Graphscape: A model for automated reasoning about visualization similarity and sequencing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 2628–2638, 2017d.

Andy Kirk. *Data visualisation : a handbook for data driven design*. SAGE, Los Angeles, 2016. ISBN 9781473912137.

Grady. Klein. *The cartoon introduction to statistics*. Hill and Wang, a Division of Farrar, Straus and Giroux, New York, first edition. edition, 2013. ISBN 0809033666.

Rex B Kline. What's wrong with statistical tests–and where we go from here. 2004.

KnightLab. Available at: http://storyline.knightlab.com/., a. [online: last accessed January 14, 2022].

KnightLab. Available at: https://timeline.knightlab.com/., b. [online: last accessed January 14, 2022].

Ha-Kyung Kong, Zhicheng Liu, and Karrie Karahalios. Frames and slants in titles of visualizations on controversial topics. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 1–12, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450356206. URL https://doi.org/10.1145/3173574.3174012.

Juuso Koponen and Jonatan Hildén. *Data visualization handbook*. Aalto korkeakoulusäätiö, 2019.

Robert Kosara and Jock Mackinlay. Storytelling: The next step for visualization. *Computer*, 46 (5):44–50, 2013. doi: 10.1109/MC.2013.36.

Zoltan Kovecses. *Metaphor: A practical introduction*. Oxford University Press, 2010.

Stephanie A. Kraft, Melissa Constantine, David Magnus, Kathryn M. Porter, Sandra Soo-Jin Lee, Michael Green, Nancy E. Kass, Benjamin S. Wilfond, and Mildred K. Cho. A randomized study of multimedia informational aids for research on medical practices: Implications for informed consent. *Clinical Trials*, 14(1):94–102, September 2016. ISSN 1740-7753. doi: 10.1177/1740774516669352.

Petera KUN, Ingrida MULDER, and Gerda KORTUEM. Data exploration for generative design research. In *Design Research Society 2018Design Research Society*, pages 1342–1345. Design Research Society, 2018.

Bum Chul Kwon, Florian Stoffel, Dominik Jäckle, Bongshin Lee, and Daniel Keim. Visjockey: Enriching data stories through orchestrated interactive visualization. In *Poster compendium of the computation+ journalism symposium*, volume 3, page 3, 2014.

Margaret L. Satterthwaite, Enrico Bertini, Oded Nov, John Emerson, Anshul Vikram Pandey, and Jeremy Boy. Data and visualization hazards. Available at: `http://visualizingrights.org/kit/hazards/index.html`, 2019. [online: last accessed January 14, 2022].

Sam Lau and Philip J Guo. Data Theater: A live programming environment for prototyping data-driven explorable explanations. In *Workshop on Live Programming (LIVE)*, 2020.

Bongshin Lee, Rubaiat Habib Kazi, and Greg Smith. Sketchstory: Telling more engaging stories with data through freeform sketching. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2416–2425, 2013. doi: 10.1109/TVCG.2013.191.

Bongshin Lee, Nathalie Henry Riche, Petra Isenberg, and Sheelagh Carpendale. More than telling a story: Transforming data into visually shared stories. *IEEE Computer Graphics and Applications*, 35(5):84–90, 2015. doi: 10.1109/MCG.2015.99.

Sukwon Lee, Sung-Hee Kim, Ya-Hsin Hung, Heidi Lam, Youn-Ah Kang, and Ji Soo Yi. How do people make sense of unfamiliar visualizations?: A grounded model of novice's information visualization sensemaking. *IEEE Transactions on Visualization and Computer Graphics*, 22 (1):499–508, 2016. doi: 10.1109/TVCG.2015.2467195.

Fritz Lekschas, Xinyi Zhou, Wei Chen, Nils Gehlenborg, Benjamin Bach, and Hanspeter Pfister. A generic framework and library for exploration of small multiples through interactive piling. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):358–368, 2021. doi: 10.1109/TVCG.2020.3028948.

W Howard Levie and Richard Lentz. Effects of text illustrations: A review of research. *ECTJ*, 30(4):195–232, 1982. doi: https://doi.org/10.1007/BF02765184.

Makayla Lewis. Sketchnotes in ux design cheat sheet. Available at: https://makaylalewis.files.wordpress.com/2018/09/ux-cambridge-2018-cheat-sheet.pdf, 2018. [online: last accessed January 14, 2022].

Endre M Lidal, Helwig Hauser, and Ivan Viola. Geological storytelling: graphically exploring and communicating geological sketches. In *Proceedings of the International Symposium on Sketch-Based Interfaces and Modeling*, pages 11–20, 2012.

J. Liem, C. Perin, and J. Wood. Structure and empathy in visual data storytelling: Evaluating their influence on attitude. volume 39, pages 277–289, 2020. doi: https://doi.org/10.1111/cgf.13980. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13980.

Huifen Lin and Tsuiping Chen. Reading authentic efl text using visualization and advance organizers in a multimedia learning environment. 2007.

Shu-Fen Lin, Huann-shyang Lin, Ling Lee, and Larry D Yore. Are science comics a good medium for science communication? the case for public learning of nanotechnology. *International Journal of Science Education, Part B*, 5(3):276–294, 2015.

William Little. *The shorter oxford english dictionary 3rd ed.* Oxford, 1956.

Zhicheng Liu, John Thompson, Alan Wilson, Mira Dontcheva, James Delorey, Sam Grigg, Bernard Kerr, and John Stasko. Data illustrator: Augmenting vector design tools with lazy data binding for expressive visualization authoring. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 1–13, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450356206. URL https://doi.org/10.1145/3173574.3173697.

Min Lu, Chufeng Wang, Joel Lanir, Nanxuan Zhao, Hanspeter Pfister, Daniel Cohen-Or, and Hui Huang. Exploring visual information flows in infographics. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, page 1–12, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450367080. URL https://doi.org/10.1145/3313831.3376263.

Patrik Lundblad and Mikael Jern. Geovisual analytics and storytelling using html5. In *2013 17th International Conference on Information Visualisation*, pages 263–271, 2013. doi: 10.1109/IV.2013.35.

Giorgia Lupi and Stefanie Posavec. *Dear data*. Chronicle books, 2016.

Giorgia Lupi and Stefanie Posavec. *Observe, Collect, Draw: A Visual Journal*. Chronicle Books, 2018.

Michael Lynch and Steve Woolgar. Representation in scientific practice. 1990.

Matt Madden. 99 ways to tell a story: Exercises in style. *New York, NY: Chamberlain Bros*, 2005.

Narges Mahyar, Sung-Hee Kim, and Bum Chul Kwon. Towards a taxonomy for evaluating user engagement in information visualization. In *Workshop on Personal Visualization: Exploring Everyday Life*, volume 3, page 2, 2015.

Adam V Maltese, Joseph A Harsh, and Dubravka Svetina. Data visualization literacy: Investigating data interpretation along the novice—expert continuum. *Journal of College Science Teaching*, 45(1):84–90, 2015.

A.D. Manning. Understanding comics: The invisible art. *IEEE Transactions on Professional Communication*, 41(1):66–69, 1998. doi: 10.1109/TPC.1998.661632.

Michele Mauri, Tommaso Elli, Giorgio Caviglia, Giorgio Uboldi, and Matteo Azzi. Rawgraphs: A visualisation platform to create open outputs. In *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter*, CHItaly '17, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450352376. doi: 10.1145/3125571.3125585. URL `https://doi.org/10.1145/3125571.3125585`.

Richard E. Mayer and Joan K. Gallini. When is an illustration worth ten thousand words? *Journal of Educational Psychology*, 82(4):715–726, 1990. ISSN 1939-2176 0022-0663. doi: 10.1037/0022-0663.82.4.715.

Scott McCloud. *Understanding comics: The invisible art*. 1993.

Scott McCloud. *Reinventing comics: How imagination and technology are revolutionizing an art form*. Harper Collins, 2000.

Scott McCloud. *Making comics*. Harper Collins, 2011.

Christopher M. McDermott and Gina Colarelli O'Connor. Managing radical innovation: An overview of emergent strategy issues. *Journal of Product Innovation Management*,

19(6):424–438, 2002. doi: https://doi.org/10.1111/1540-5885.1960424. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1540-5885.1960424.

Richard McGuire. *Here*. Raw Books & Graphics, 1989.

S. McKenna, N. Henry Riche, B. Lee, J. Boy, and M. Meyer. Visual narrative flow: Exploring factors shaping data visualization story reading experiences. volume 36, pages 377–387, 2017a. doi: https://doi.org/10.1111/cgf.13195. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13195.

Sean McKenna, Dominika Mazur, James Agutter, and Miriah Meyer. Design activity framework for visualization design. *IEEE Transactions on Visualization and Computer Graphics*, 20(12): 2191–2200, 2014. doi: 10.1109/TVCG.2014.2346331.

Sean McKenna, Alexander Lex, and Miriah Meyer. Worksheets for guiding novices through the visualization design process. *arXiv preprint arXiv:1709.05723*, 2017b.

Michael C. Medlock, Dennis Wixon, Mick McGee, and Dan Welsh. The rapid iterative test and evaluation method: Better products in less time. In Randolph G. Bias and Deborah J. Mayhew, editors, *Cost-Justifying Usability (Second Edition)*, Interactive Technologies, pages 489–517. Morgan Kaufmann, San Francisco, second edition edition, 2005. ISBN 978-0-12-095811-5. doi: https://doi.org/10.1016/B978-012095811-5/50017-1. URL https://www.sciencedirect.com/science/article/pii/B9780120958115500171.

Andrei Molotiu. *Abstract comics: The anthology: 1967-2009*. Fantagraphics Books, 2009.

Richard Monastersky and Nick Sousanis. The fragile framework. *Nature*, 527(7579):1, 2015.

Mark Monmonier. *How to lie with maps*. University of Chicago Press, 2018.

Randall Munroe. Xkcd. Available at: https://xkcd.com/. [online: last accessed January 14, 2022].

Randall Munroe. *Thing explainer: complicated stuff in simple words*. Hachette UK, 2015.

Tamara Munzner. A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):921–928, 2009. doi: 10.1109/TVCG.2009.111.

Tamara Munzner. *Visualization analysis and design*. AK Peters/CRC Press, 2014.

Aquiles Negrete and Cecilia Lartigue. Learning from education to communicate science as a good story. *Endeavour*, 28(3):120–124, 2004. ISSN 0160-9327. doi: https://doi.org/10.1016/j.endeavour.2004.07.003. URL https://www.sciencedirect.com/science/article/pii/S0160932704000766.

Deborah Nolan and Jamis Perrett. Teaching and learning data visualization: Ideas and assignments. *The American Statistician*, 70(3):260–269, 2016.

Heather L. O'Brien and Elaine G. Toms. The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology*, 61 (1):50–69, 2010. doi: https://doi.org/10.1002/asi.21229. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.21229.

Michael O'Donnell. Why doctors don't read research papers: scientific papers are not written to disseminate information. *BMJ: British Medical Journal*, 330(7485):256, 2005.

Chris Olah and Shan Carter. Research debt. *Distill*, 2(3):e5, 2017.

Allan Paivio. *Mental representations: A dual coding approach*. Oxford University Press, 1990.

Anshul Vikram Pandey, Katharina Rall, Margaret L. Satterthwaite, Oded Nov, and Enrico Bertini. How deceptive are deceptive visualizations? an empirical analysis of common distortion techniques. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, page 1469–1478, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450331456. doi: 10.1145/2702123.2702608. URL https://doi.org/10.1145/2702123.2702608.

Luc Pauwels. *Visual cultures of science: rethinking representational practices in knowledge building and science communication*. UPNE, 2006.

Roxy Peck, Chris Olsen, and Jay L Devore. *Introduction to statistics and data analysis*. Nelson Education, 2011.

C.L. Philip Chen and Chun-Yang Zhang. Data-intensive applications, challenges, techniques and technologies: A survey on big data. *Information Sciences*, 275:314–347, 2014. ISSN 0020-0255. doi: https://doi.org/10.1016/j.ins.2014.01.015. URL https://www.sciencedirect.com/science/article/pii/S0020025514000346.

Zachary Pousman, John Stasko, and Michael Mateas. Casual information visualization: Depictions of data in everyday life. *IEEE Transactions on Visualization and Computer Graphics*, 13 (6):1145–1152, 2007. doi: 10.1109/TVCG.2007.70541.

Zening Qu and Jessica Hullman. Keeping multiple views consistent: Constraints, validations, and exceptions in visualization authoring. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):468–477, 2018. doi: 10.1109/TVCG.2017.2744198.

Ibn al Rabin. *Splendeurs & misères du verbe*. L'Association, 2012.

Donghao Ren, Matthew Brehmer, Bongshin Lee, Tobias Höllerer, and Eun Kyoung Choe. Chartaccent: Annotation for data-driven storytelling. In *2017 IEEE Pacific Visualization Symposium (PacificVis)*, pages 230–239, 2017. doi: 10.1109/PACIFICVIS.2017.8031599.

Donghao Ren, Bongshin Lee, and Matthew Brehmer. Charticulator: Interactive construction of bespoke chart layouts. *IEEE Transactions on Visualization and Computer Graphics*, 25(1): 789–799, 2019. doi: 10.1109/TVCG.2018.2865158.

Sandra Rendgen. *The Minard System: The Complete Statistical Graphics of Charles-Joseph Minard*. Chronicle Books, 2018.

Sandra Rendgen, Julius Wiedemann, Paolo Ciuccarelli, Richard Saul Wurman, Simon Rogers, and Nigel Holmes. *Information graphics*. Taschen Cologne, Germany, 2012.

Severino Ribecca. The data visualization catalogue. Available at: `https://datavizcatalogue.com`, 2019. [online: last accessed January 14, 2022].

Nathalie Henry Riche, Christophe Hurter, Nicholas Diakopoulos, and Sheelagh Carpendale. *Data-driven storytelling*. CRC Press, 2018.

Jacob Ritchie, Daniel Wigdor, and Fanny Chevalier. A lie reveals the truth: Quasimodes for task-aligned data presentation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, page 1–13, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450359702. URL `https://doi.org/10.1145/3290605.3300423`.

Jonathan C. Roberts, Chris Headleand, and Panagiotis D. Ritsos. Sketching designs using the five design-sheet methodology. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):419–428, 2016. doi: 10.1109/TVCG.2015.2467271.

Simon Rogers and Mark Mccormick. Available at: http://image.guardian.co.uk/sys-files/Guardian/documents/2007/12/17/CARBON_ATLAS.pdf. Original website not available anymore., 2010. [online: last accessed January 14, 2022].

Hugo Romat, Nathalie Henry Riche, Christophe Hurter, Steven Drucker, Fereshteh Amini, and Ken Hinckley. Dear pictograph: Investigating the role of personalization and immersion for consuming and enjoying visualizations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, page 1–13, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450367080. URL https://doi.org/10.1145/3313831.3376348.

Gertjan Rop, Anne Schüler, Peter P.J.L. Verkoeijen, Katharina Scheiter, and Tamara Van Gog. The effect of layout and pacing on learning from diagrams with unnecessary text. *Applied Cognitive Psychology*, 32(5):610–621, 2018. doi: https://doi.org/10.1002/acp.3445. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/acp.3445.

Hans Rosling. Gapminder. Available at: https://www.gapminder.org/, a. [online: last accessed January 14, 2022].

Hans Rosling. Myths about the so-called "developing world". Available at: https://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen#t-230391, b. [online: last accessed January 14, 2022].

Hannah R. Rothstein, Alexander J. Sutton, and Michael Borenstein. Publication bias in meta-analysis. pages 1–7, 2005. doi: https://doi.org/10.1002/0470870168.ch1. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/0470870168.ch1.

Hana Roš and Matteo Farinella. *Neurocomic*. Neurocomic. Nobrow, 2013. ISBN 9788817075794.

Sherry Ruan, Jacob O. Wobbrock, Kenny Liou, Andrew Ng, and James A. Landay. Comparing speech and keyboard text entry for short messages in two languages on touchscreen phones. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(4), jan 2018. doi: 10.1145/3161187. URL https://doi.org/10.1145/3161187.

Puripant Ruchikachorn and Klaus Mueller. Learning visualizations by analogy: Promoting visual literacy through visualization morphing. *IEEE Transactions on Visualization and Computer Graphics*, 21(9):1028–1044, 2015. doi: 10.1109/TVCG.2015.2413786.

Adam Rule, Aurélien Tabard, and James D Hollan. Exploration and explanation in computational notebooks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2018.

Lindy Ryan. *Visual Data Storytelling with Tableau: Story Points, Telling Compelling Data Narratives*. Addison-Wesley Professional, 2018.

Bahador Saket, Alex Endert, and John Stasko. Beyond usability and performance: A review of user experience-focused evaluations in visualization. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, BELIV '16, page 133–142, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450348188. doi: 10.1145/2993901.2993903. URL https://doi.org/10.1145/2993901.2993903.

Arvind Satyanarayan and Jeffrey Heer. Authoring narrative visualizations with ellipsis. volume 33, pages 361–370, 2014. doi: https://doi.org/10.1111/cgf.12392. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.12392.

Arvind Satyanarayan, Dominik Moritz, Kanit Wongsuphasawat, and Jeffrey Heer. Vega-lite: A grammar of interactive graphics. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):341–350, 2017. doi: 10.1109/TVCG.2016.2599030.

Jeffrey D Scargle. Publication bias (the "file-drawer problem") in scientific inference. *arXiv preprint physics/9909033*, 1999.

Milo Schield. Information literacy, statistical literacy and data literacy. In *IASSIST QUARTERLY (IQ)*. Citeseer, 2004.

Jon Schwabish. The graphic continuum flash cards. Available at: https://policyviz.com/2016/10/31/introducing-graphic-continuum-flash-cards/, 2016. [online: last accessed January 14, 2022].

Jon Schwabish. Core principles of data visualization cheatsheet. Available at: https://policyviz.com/product/core-principles-of-data-visualization-cheatsheet/, 2019. [online: last accessed January 14, 2022].

Edward Segel and Jeffrey Heer. Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1139–1148, 2010. doi: 10.1109/TVCG.2010.179.

Priti Shah and James Hoeffner. Review of graph comprehension research: Implications for instruction. *Educational psychology review*, 14(1):47–69, 2002. doi: https://doi.org/10.1023/A:1013180410169.

Danqing Shi, Xinyue Xu, Fuling Sun, Yang Shi, and Nan Cao. Calliope: Automatic visual data story generation from a spreadsheet. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):453–463, 2021. doi: 10.1109/TVCG.2020.3030403.

Ben Shneiderman. Creativity support tools: Accelerating discovery and innovation. *Commun. ACM*, 50(12):20–32, dec 2007. ISSN 0001-0782. doi: 10.1145/1323688.1323689. URL https://doi.org/10.1145/1323688.1323689.

Corry Shores. "ragged time" in intra-panel comics rhythms. *The Comics Grid: Journal of Comics Scholarship*, 6, 2016.

I. Simera, D. Moher, J. Hoey, K. F. Schulz, and D. G. Altman. A catalogue of reporting guidelines for health research. *European Journal of Clinical Investigation*, 40(1):35–53, 2010. doi: https://doi.org/10.1111/j.1365-2362.2009.02234.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2362.2009.02234.x.

Ian Simon, Dan Morris, and Sumit Basu. Mysong: Automatic accompaniment generation for vocal melodies. In *Proceedings of the 2008 CHI Conference on Human Factors in Computing Systems*, CHI '08, page 725–734, New York, NY, USA, 2008. Association for Computing Machinery. ISBN 9781605580111. doi: 10.1145/1357054.1357169. URL https://doi.org/10.1145/1357054.1357169.

Alan Smith, Chris Campbell, Ian Bott, Liz Faunce, Graham Parrish, Billy Ehrenberg-Shannon, Paul McCallum, and Martin Stabe. Chartdoctor. https://github.com/ft-interactive/chart-doctor/blob/master/visual-vocabulary/Visual-vocabulary.pdf, 2019. [online: last accessed January 14, 2022].

Woollcott Smith and Larry Gonick. *The cartoon guide to statistics*. Collins Reference, 2005.

Edirlei Soares de Lima, Bruno Feijó, Antonio L. Furtado, Simone Diniz Junqueira Barbosa, Cesar T. Pozzer, and Angelo E. M. Ciarlini. Non-branching interactive comics. In Dennis Reidsma, Haruhiro Katayose, and Anton Nijholt, editors, *Advances in Computer Entertainment*, pages 230–245, Cham, 2013. Springer. ISBN 978-3-319-03161-3.

Yang Song and David Thuente. A quantitative case study in engineering of the efficacy of quality cheat-sheets. In *2015 IEEE Frontiers in Education Conference (FIE)*, pages 1–7, 2015. doi: 10.1109/FIE.2015.7344082.

Madeleine Sorapure. Text, image, data, interaction: Understanding information visualization. *Computers and Composition*, 54:102519, 2019. ISSN 8755-4615. doi: https://doi.org/10.1016/j.compcom.2019.102519. URL https://www.sciencedirect.com/science/article/pii/S875546151830032X.

Nick Sousanis. *Unflattening*. Harvard University Press, Cambridge, Massachusetts, April 2015. ISBN 978-0-674-74443-1.

Amy N Spiegel, Julia McQuillan, Peter Halpin, Camillia Matuk, and Judy Diamond. Engaging teenagers with science through comics. *Research in science education*, 43(6):2309–2326, 2013. doi: https://doi.org/10.1007/s11165-013-9358-x.

CD Stolper, B Lee, N Henry Riche, J Stasko, C Hurter, N Diakopoulos, and S Carpendale. Data-driven storytelling techniques: Analysis of a curated collection of visual stories. *Data-Driven Storytelling*, pages 85–105, 2018.

Charles D Stolper, Bongshin Lee, Nathalie Henry Riche, and John Stasko. Emerging and recurring data-driven storytelling techniques: Analysis of a curated collection of recent stories. 2016.

Hendrik Strobelt, Daniela Oelke, Christian Rohrdantz, Andreas Stoffel, Daniel A. Keim, and Oliver Deussen. Document cards: A top trumps visualization for documents. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1145–1152, 2009. doi: 10.1109/TVCG.2009.139.

Donna F Stroup, Jesse A Berlin, Sally C Morton, Ingram Olkin, G David Williamson, Drummond Rennie, David Moher, Betsy J Becker, Theresa Ann Sipe, Stephen B Thacker, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *Jama*, 283(15): 2008–2012, 2000.

Nicole Sultanum, Michael Brudno, Daniel Wigdor, and Fanny Chevalier. More text please! understanding and supporting the use of visualization for clinical text overview. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 1–13, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450356206. URL https://doi.org/10.1145/3173574.3173996.

Nicole Sultanum, Fanny Chevalier, Zoya Bylinskii, and Zhicheng Liu. Leveraging text-chart links to support authoring of data-driven articles with vizflow. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380966. doi: 10.1145/3411764.3445354. URL https://doi.org/10.1145/3411764.3445354.

Tableau. Gapminder. Available at: https://public.tableau.com/en-us/gallery/?tab=viz-of-the-day&type=viz-of-the-day. [online: last accessed January 14, 2022].

Mohammad Tahaei, Kami Vaniea, and Naomi Saphra. Understanding privacy-related questions on stack overflow. page 1–14, 2020. URL https://doi.org/10.1145/3313831.3376768.

Shin Takahashi et al. *The manga guide to statistics*. No Starch Press, 2008.

Mico Tatalovic. Science comics as tools for science education and communication: a brief, exploratory study. *Journal of Science Communication*, 8(4):A02, 2009.

The New York Times. You draw it: How family income predicts children's college chances. Available at: https://www.nytimes.com/interactive/2015/05/28/upshot/you-draw-it-how-family-income-affects-childrens-college-chances.html, 2015. [online: last accessed January 14, 2022].

The New York Times. You draw it: Just how bad is the drug overdose epidemic. Available at: https://www.nytimes.com/interactive/2017/04/14/upshot/drug-overdose-epidemic-you-draw-it.html, 2017. [online: last accessed January 14, 2022].

The Washington Post. Why outbreaks like coronavirus spread exponentially, and how to 'flatten the curve'. Available at: https://www.washingtonpost.com/graphics/2020/world/corona-simulator/, 2020. [online: last accessed January 14, 2022].

Alice Thudt, Uta Hinrichs, and Sheelagh Carpendale. The bohemian bookshelf supporting serendipitous discoveries through visualization. Technical report, University of Calgary, 2011.

Hiroaki Tobita. Digestmanga: interactive movie summarizing through comic visualization. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, pages 3751–3756. 2010.

Hiroaki Tobita. Comic-crowd: interactive comic creation that supports multiple storylines, visualizations, and platforms. In *Proc. Conference on Mobile and Ubiquitous Multimedia*, pages 163–172, 2015.

Hiroaki Tobita and Kaoru Shibasaki. Enformanga: Interactive comic creation using drag-and-drop and deformation. In *2009 11th IEEE International Symposium on Multimedia*, pages 269–274, 2009. doi: 10.1109/ISM.2009.126.

Chao Tong, Richard Roberts, Rita Borgo, Sean Walton, Robert S Laramee, Kodzo Wegba, Aidong Lu, Yun Wang, Huamin Qu, Qiong Luo, et al. Storytelling and visualization: An extended survey. *Information*, 9(3):65, 2018.

Anne M. Treisman and Garry Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136, 1980. ISSN 0010-0285. doi: https://doi.org/10.1016/0010-0285(80)90005-5. URL https://www.sciencedirect.com/science/article/pii/0010028580900055.

Edward R Tufte. The visual display of quantitative information. *The Journal for Healthcare Quality (JHQ)*, 7(3):15, 1985.

John W Tukey et al. *Exploratory data analysis*, volume 2. Reading, Mass., 1977.

BARBARA TVERSKY, JULIE BAUER MORRISON, and MIREILLE BETRANCOURT. Animation: can it facilitate? *International Journal of Human-Computer Studies*, 57(4): 247–262, 2002. ISSN 1071-5819. doi: https://doi.org/10.1006/ijhc.2002.1017. URL https://www.sciencedirect.com/science/article/pii/S1071581902910177.

Shingo Uchihashi, Jonathan Foote, Andreas Girgensohn, and John Boreczky. Video manga: Generating semantically meaningful video summaries. In *Proceedings of the Seventh ACM International Conference on Multimedia (Part 1)*, MULTIMEDIA '99, page 383–392, New York, NY, USA, 1999. Association for Computing Machinery. ISBN 1581131518. doi: 10.1145/319463.319654. URL https://doi.org/10.1145/319463.319654.

J Valero, J Català, and BE Marín. An approach to the taxonomy of data visualisation. *Revista Latina de Comunicación Social*, 69:486–507, 2014.

Veen van Fjodor. The neural network zoo. Available at: https://www.asimovinstitute.org/neural-network-zoo/, 2016. [online: last accessed January 14, 2022].

Bret Victor. Explorable explanations. Available at: http://worrydream.com/ExplorableExplanations/, 2011a. [online: last accessed January 14, 2022].

Bret Victor. Scientific communication as sequential art. Available at: http://worrydream.com/ScientificCommunicationAsSequentialArt/, 2011b. [online: last accessed January 14, 2022].

Bret Victor. Tangle: explorable explanations made easy. Available at: http://worrydream.com/Tangle, 2013. [online: last accessed January 14, 2022].

Fernanda B Viegas and Martin Wattenberg. Communication-minded visualization: A call to action. *IBM Systems Journal*, 45(4):801, 2006.

Yoav Wachsman et al. Should cheat sheets be used as study aids in economics tests. *Economics Bulletin*, 1(1):1–10, 2002.

Qianwen Wang, Zhen Li, Siwei Fu, Weiwei Cui, and Huamin Qu. Narvis: Authoring narrative slideshows for introducing data visualization designs. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):779–788, 2019a. doi: 10.1109/TVCG.2018.2865232.

Zezhong Wang, Harvey Dingwall, and Benjamin Bach. Teaching data visualization and storytelling with data comic workshops. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA '19, page 1–9, New York, NY, USA, 2019b. Association for Computing Machinery. ISBN 9781450359719. doi: 10.1145/3290607.3299043. URL https://doi.org/10.1145/3290607.3299043.

Zezhong Wang, Shunming Wang, Matteo Farinella, Dave Murray-Rust, Nathalie Henry Riche, and Benjamin Bach. Comparing effectiveness and engagement of data comics and infographics. page 1–12, 2019c. URL https://doi.org/10.1145/3290605.3300483.

Zezhong Wang, Lovisa Sundin, Dave Murray-Rust, and Benjamin Bach. Cheat sheets for data visualization techniques. 2020.

Zezhong Wang, Jacob Ritchie, Jingtao Zhou, Fanny Chevalier, and Benjamin Bach. Data comics for reporting controlled user studies in human-computer interaction. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):967–977, 2021. doi: 10.1109/TVCG.2020.3030433.

Zezhong Wang, Hugo Romat, Fanny Chevalier, Nathalie Henry Riche, Dave Murray-Rust, and Benjamin Bach. Interactive data comics. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):944–954, 2022. doi: 10.1109/TVCG.2021.3114849.

Chris Ware, Marnie Ware, and Clara Ware. *Building stories*. Pantheon Books New York, 2012.

M. Wattenberg. Baby names, visualization, and social data analysis. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pages 1–7, 2005. doi: 10.1109/INFVIS.2005.1532122.

Hadley Wickham and Lisa Stryjewski. 40 years of boxplots. *Am. Statistician*, 2011.

L. Wilkinson, A. Anand, and R. Grossman. Graph-theoretic scagnostics. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pages 157–164, 2005. doi: 10.1109/INFVIS.2005.1532142.

Leland Wilkinson. The grammar of graphics. In James E. Gentle, Wolfgang Karl Härdle, and Yuichi Mori, editors, *Handbook of Computational Statistics: Concepts and Methods*, pages 375–414. Springer, Berlin, Heidelberg, 2012. ISBN 978-3-642-21551-3. doi: 10.1007/978-3-642-21551-3_13. URL https://doi.org/10.1007/978-3-642-21551-3_13.

Michael Wohlfart. Story telling aspects in medical applications. In *Central European Seminar on Computer Graphics*. Citeseer, 2006.

Jo Wood, Petra Isenberg, Tobias Isenberg, Jason Dykes, Nadia Boukhelifa, and Aidan Slingsby. Sketchy rendering for information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2749–2758, 2012. doi: 10.1109/TVCG.2012.262.

Haijun Xia, Nathalie Henry Riche, Fanny Chevalier, Bruno De Araujo, and Daniel Wigdor. Dataink: Direct and creative data-oriented drawing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2018.

Gene Yang. Graphic novels in the classroom. *Language Arts*, 85(3):185, 2008.

Shawn C. Yarnes. Introduction to randomization and layout. Available at: https://pbgworks.org/index.php?q=node/1534, 2013. [online: last accessed January 14, 2022].

Stephanie Yee and Tony Chu. A visual introduction to machine learning. Available at: http://www.r2d3.us/visual-intro-to-machine-learning-part-1/, 2015. [online: last accessed January 14, 2022].

Polle T. Zellweger, Bay-Wei Chang, and Jock D. Mackinlay. Fluid links for informed and incremental link transitions. In *Proceedings of the Ninth ACM Conference on Hypertext and Hypermedia : Links, Objects, Time and Space—Structure in Hypermedia Systems: Links, Objects, Time and Space—Structure in Hypermedia Systems*, HYPERTEXT '98, page 50–57,

New York, NY, USA, 1998. Association for Computing Machinery. ISBN 0897919726. doi: 10.1145/276627.276633. URL https://doi.org/10.1145/276627.276633.

Polle T. Zellweger, Anne Mangen, and Paula Newman. Reading and writing fluid hypertext narratives. In *Proceedings of the Thirteenth ACM Conference on Hypertext and Hypermedia*, HYPERTEXT '02, page 45–54, New York, NY, USA, 2002. Association for Computing Machinery. ISBN 1581134770. doi: 10.1145/513338.513353. URL https://doi.org/10.1145/513338.513353.

Jiayi Eris Zhang, Nicole Sultanum, Anastasia Bezerianos, and Fanny Chevalier. Dataquilt: Extracting visual elements from images to craft pictorial visualizations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, page 1–13, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450367080. URL https://doi.org/10.1145/3313831.3376172.

Jian Zhao, Shenyu Xu, Senthil Chandrasegaran, Chris Bryan, Fan Du, Aditi Mishra, Xin Qian, Yiran Li, and Kwan-Liu Ma. Chartstory: Automated partitioning, layout, and captioning of charts into comic-style narratives. *arXiv preprint arXiv:2103.03996*, 2021.

Zhenpeng Zhao, William Benjamin, Niklas Elmqvist, and Karthik Ramani. Sketcholution: Interaction histories for sketching. *International Journal of Human-Computer Studies*, 82: 11–20, 2015a. ISSN 1071-5819. doi: https://doi.org/10.1016/j.ijhcs.2015.04.003. URL https://www.sciencedirect.com/science/article/pii/S1071581915000725.

Zhenpeng Zhao, Rachael Marr, and Niklas Elmqvist. Data comics: Sequential art for data-driven storytelling. *tech. report*, 2015b.

Zhenpeng Zhao, Rachael Marr, Jason Shaffer, and Niklas Elmqvist. Understanding partitioning and sequence in data-driven storytelling. In Natalie Greene Taylor, Caitlin Christian-Lamb, Michelle H. Martin, and Bonnie Nardi, editors, *Information in Contemporary Society*, pages 327–338, Cham, 2019. Springer. ISBN 978-3-030-15742-5.

Qiyu Zhi, Alvitta Ottley, and Ronald Metoyer. Linking and layout: Exploring the integration of text and visualization in storytelling. In *Computer Graphics Forum*, volume 38, pages 675–685, 2019.

Caroline Ziemkiewicz and Robert Kosara. Preconceptions and individual differences in understanding visual metaphors. volume 28, pages 911–918, 2009. doi: https://doi.org/10.

1111/j.1467-8659.2009.01442.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8659.2009.01442.x.